

NBER WORKING PAPER SERIES

ARE THERE TOO MANY FARMS IN THE WORLD? LABOR-MARKET TRANSACTION
COSTS, MACHINE CAPACITIES AND OPTIMAL FARM SIZE

Andrew D. Foster
Mark R. Rosenzweig

Working Paper 23909
<http://www.nber.org/papers/w23909>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2017, Revised March 2021

This research supported in part through NIH Grant P2C HD041020. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Andrew D. Foster and Mark R. Rosenzweig. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size

Andrew D. Foster and Mark R. Rosenzweig

NBER Working Paper No. 23909

October 2017, Revised March 2021

JEL No. O13

ABSTRACT

This paper seeks to explain the U-shaped relationship between farm productivity and farm scale - the initial fall in productivity as farm size increases from its lowest levels and the continuous upward trajectory as scale increases after a threshold - observed across the world and in low-income countries. We show that the existence of labor-market transaction costs can explain why the smallest farms are most efficient, slightly larger farms least efficient and larger farms as efficient as the smallest farms. We show that to explain the rising upper tail of the U characteristic of high-income countries requires there be economies of scale in the ability of machines to accomplish tasks at lower costs at greater operational scales. Using data from the India ICRISAT VLS panel survey we find evidence consistent with these conditions, suggesting that there are too many farms, at scales insufficient to exploit locally-available equipment-capacity scale-economies.

Andrew D. Foster
Department of Economics and Community Health
Brown University
64 Waterman Street
Providence, RI 02912
and NBER
afoster@brown.edu

Mark R. Rosenzweig
Department of Economics
Yale University
Box 208269
New Haven, CT 06520
and NBER
mark.rosenzweig@yale.edu

“... the most dominant as well as the most intractable feature of our agrarian economy is the small size of the holding occupied by the vast majority of the cultivators. No effective solution of the problem of improved production and the crushing burden of poverty can be found until we devise a system in which the unit of agricultural organisation will not ordinarily be below the minimum unit.” United Provinces Zamindari Abolition Committee, 1948, p.501.

This paper revisits the issue of the relationship between operation scale and productivity in agriculture. The research is motivated by three stylized facts characterizing world agriculture. First, farming in low-income countries is small scale while farming in developed countries is large scale. Figure 1 displays the proportions of operational holdings of farms that are below 10 acres across a sample of developed and developing countries for which reliable data are available on the size distribution of farms. As can be seen, only 10% or less of farms are below 10 acres in the United States and Canada, while for the three most populous low-income countries - China, India, and Indonesia - at least 80% of farms are below 10 acres. In major economies in Africa too, as seen in the figure, only a small proportion of farms are above 10 acres. Note that these figures may underestimate the extent of small-scale farming in such countries to the extent that the landholdings of a farm are fragmented into spatially-separated plots.

The second stylized fact is that the productivity of developed-country agriculture is substantially higher than it is in low-income countries. For example, as shown in Figure 2, soybean yields are four times higher in the United States, where farm scale is high, than they are in Indonesia, India and the Philippines, where farms are small, and three times higher in Canada. This figure also illustrates, however, why output per acre is insufficient to gauge productivity - China appears to be an outlier in that its yields are twice as high as those in the other three low-income countries in the figure, despite its similarity in operational scale. However, this is misleading, as the fertilizer-intensity in China, as seen in Figure 3, is 2.7 to 3.5 times higher than that in Indonesia, India, and the

Philippines and five times higher than that in the United States.¹ Assessing productivity requires attention to input use and its cost.

An implication of any positive causal relationship between production scale and agricultural productivity implied by the differences in scale and productivity across countries is that there are too many farms in the world, especially in low-income countries. It implies that enlarging the size of farms via consolidation would increase overall agricultural output, with an accompanying substantial reduction in the amount of poverty and employment in agriculture. This was the conclusion reached by the United Provinces Zamindari Committee, charged in 1946 with recommending the redistribution of the large landholdings of the Zamindari in the United provinces of India, in its 1948 report. It suggested a specific minimal scale of operation of 10 acres to boost productivity, based on the principal mode of motive power in India at the time - a bullock pair. They then concluded that consolidating landholdings to meet that minimum size would entail a reduction in farms of 66%.

Of course, differences between developed-country and developing-country agriculture are due to more than scale. The best evidence on scale relevant for a low-income country would come from a single country, based on farms in the same institutional environment, the same markets and facing the same technology frontier. When farm scale and farm productivity are examined within a country, however, we observe the third stylized fact: there is an almost universal inverse relationship between farm or plot size and productivity within developing countries, over the span of plot and farm sizes observed in those countries, while continuous increasing returns to scale are observed among the larger farms in developed countries (e.g. Paul *et al.*, 2004).

While most of the literature documenting the inverse relationship in low-income countries is

¹The evidence suggests that fertilizer is over-used in China, with the marginal return on a dollar of fertilizer less than 7 cents (Huang *et al.*, 2008). The reasons for the high fertilizer intensity in China are the large fertilizer subsidies and governmental resource assessments based on crop yields rather than net returns.

based on data from India, the Philippines and Latin America (e.g, Schultz, 1964; Hayami and Otsuka, 1993; Binswanger *et al.*, 1995; Hazell, 2011; Vollrath, 2007; Kagin *et al.*, 2015), recently-available, representative data sets reveal this same inverse relationship, based on well-measured land areas, holds almost universally. Figures 4A-4D display the relationships between farm area and output per acre in China, in Nigeria, in Mexico and in Bangladesh.² In all four of these data sets based on per-acre yields, the very smallest farms are substantially more productive. And, as can also be seen, the span of farm sizes, based on representative data, is quite limited. The existing descriptive evidence on scale and farm productivity from data describing farming in low-income countries thus does not support the notion there are too many farms, though it does indicate that land is mis-allocated given heterogeneity in productivity by size.

There is a large literature focused on low-income countries that has also attempted to address the puzzle of why the smallest farms are most productive, without little consensus. There is general agreement that the inverse relationship is not spurious - specifically, not due to a correlation of land quality and farm size (e.g., Carter, 1984; Bhalla and Roy, 1988; Benjamin, 1995; Barrett *et al.*, 2010) and/or measurement error that is correlated with scale (e.g., Ali and Deininger, 2014; Larson *et al.*, 2013; Carletto *et al.*, 2013). However, a general shortcoming of this literature is that it may be addressing the wrong puzzle. Given the global pattern of farm productivity, the puzzle that requires explanation is why there is a U-shape relationship between farm productivity and scale - why the smallest farms, which dominate low-income countries, are more productive than somewhat less small farms there and why in the developed world the large-scale farms are not only more productive on average, but productivity increases with scale.³

²The data sets used to create these graphs are the Integrated Agricultural Productivity Project 2013 (Bangladesh), the China Living Standards Survey 1995/1997, the Mexico Family Life Survey 2002., and the LSMS-Integrated Surveys on Agriculture General Household Survey Panel for 2015/16 (Nigeria).

³Seen from this global perspective, some of explanations for the inverse relationship observed in low-income countries are at best incomplete. For example, the idea that farms exclusively managed and worked on

In this paper we seek to explain the U-shaped relationship between farm productivity and farm scale with a model that incorporates labor-market transaction costs and scale economies in machine capacities. We provide tests of some of the implications of the model, and quantify a parameterized version of the model based on estimated and calibrated structural parameters to show the model is capable of yielding the U-shape productivity relationship, even if the production technology is CRS. We then use the estimates to identify the magnitudes of labor transaction costs, to estimate the optimal farm scale given existing machine technology in India, and to carry out a counterfactual land consolidation, embedding the model in a general equilibrium framework, in which all farms are at the optimal scale. The result of the counterfactual is both an increase in total output from the same total cultivated land as well as a substantial increase in output per laborer. The exercise thus enables us to quantify the number of surplus farmers and farm laborers in India and the loss in incomes associated with the existing distribution of landholdings.

We highlight transaction costs in the labor market because they are especially important in agriculture. Agricultural operations, at any scale or level of technology, are sequential, intermittent and their timing is based on unpredictable weather events - labor is thus principally hired on a daily basis, with employers and workers seeking matches at high frequency. Moreover, the amount of work needed on a given day may vary, so there is daily variation in worker hours. We show that the existence of fixed transaction costs, to the extent they are born by farmers, makes farmers at the margin at which hiring labor would be productive on net (all family labor fully utilized) reluctant to hire labor. And, if labor is hired at all, average unit labor costs will vary by operational scale because larger scale entails more intensive use of labor. The result is a U-shape in which the smallest farms

by owner-operators and their families, which characterizes the smallest farms, have an advantage, because of superior incentives, lower supervision costs, and lower unit-labor costs (Yotopoulos and Lau, 1973; Carter and Wiebe, 1990; Binswanger-Mkhize, *et al.*, 2009; Hazel *et al.*, 2010;) while true, cannot explain why corporate farms, which are large scale, are even more productive. A common finding in this literature is that the smallest farms use labor more intensively than larger farms, which would generate higher per-acre output but not necessarily higher productivity accounting for input costs, but the reasons for this are not settled.

are most efficient in their use of labor, slightly larger farms least efficient and larger farms as efficient as the smallest farms because the share of transaction costs in total labor expenditures are smallest.⁴⁵

We also show that fixed transaction costs can explain another feature of farming in low-income countries, the high share of operations that are autarchic, in which no family members work off the farm and the farm labor force includes no hired labor. Based on data describing both the off-farm labor of household members and the composition of labor working on the farm, the survey data from Nigeria indicates that 36.2% of planting operations are autarchic, the data from China that 17% of farms operated all operations over the entire agricultural under autarchy, and the data from the India that we use in this research, described below, which describe input use for every operation, indicates that 34% of operations are autarchic. All three data sets indicate that autarchic farming is not concentrated among the smallest farms but among intermediate-sized farms, as depicted in Figures 5A-5C, which show the fraction of autarchic farms or operations by land size for the three countries. We show that in such operations, labor is substantially underutilized relative to labor use on smaller and larger farms.

The existence of fixed labor market transaction cost can explain both the inverted-U in autarchic farming and the U-shape in land productivity, but it cannot account for the higher productivity of larger farms compared to the smallest farms, positive scale economies that continue

⁴Allen (1988) shows that one of the reasons that larger farms were more productive than smaller farms in 18th century England, when mechanization was not a major factor, was that larger farms could hire labor crews. If multiple workers are needed at the same time then hiring a worker team can reduce per-worker search costs for the employer. We test for, but do not find, evidence of this form of scale in the villages we study

⁵Foster and Rosenzweig (2011) highlight the additional supervision costs associated with using hired labor as an explanation for labor under-utilization at farm scales above the smallest. But supervision costs cannot explain why above a threshold larger farms become more efficient in the absence of labor-substituting mechanization since it is not likely such costs diminish as the amount of hired work increases.

at higher scales.⁶ For this, we focus on scale economies in machinery capacity. There is ample evidence that agricultural machinery saves on labor costs (Hornbeck and Naidu, 2014; Davis, 2016) and that mechanization is more likely on larger farms (Foster and Rosenzweig, 2011). If there is a minimum farm scale at which mechanized equipment can be used, then only larger farms can exploit this substitution to avoid the additional costs of hiring labor.⁷ But, again, a single threshold cannot explain the continuing rise in productivity with scale. We show that the upper tail of the U can be explained by introducing economies of scale in the capacity of machines, where higher capacity machines accomplish tasks at lower costs but are only suited to greater operational scale. There are two conditions that must be met: effective machine capacity can only be increased at larger scale and the pricing of capacity must be non-linear. We address the question of whether these conditions are met within a low-income country.

We are able to examine and quantify the role of transactions costs and machine capacity scale economies as major factors accounting for the U-shape relationship between scale and productivity within a low-income country because of the existence of unique data from the India ICRISAT VLS panel survey. One key advantage of the ICRISAT survey relative to many other representative surveys is that it seeks to balance the sample by landholding size, rather than, say

⁶Some studies have suggested that access to capital and a greater ability to insure against risk may explain why larger farms may be more productive than smaller farms. However, we show that the U-shaped relationship between scale and productivity holds across plots for the same farmer, which effectively holds constant the farmer's ability to take risk, finance capital and make better allocative decisions. We thus abstract from these considerations, but this does not imply they are not important determinants of agricultural productivity in low-income countries.

⁷The idea that there are physical constraints associated with the size of plots inhibiting mechanization is well known. Bivar (2010), in her study of the French government- and union-led agricultural consolidation program initiated in the early 1950's - motivated by the potential productivity-enhancing effects of mechanization - cites documents written by the French Agriculture ministry that suggest for a tractor to be able to turn around, a minimum plot size of 1.5-2 hectares is required. Of course, there are more mechanized options today that require a smaller minimum scale, but these may have reduced performance, which is one of the key hypotheses we test here.

population.⁸ As a consequence of this sampling frame, larger farms are over-sampled and we are able to examine both small and larger farms in a common environment. The data set thus contains the missing link between low-income country agriculture and developed-country agriculture - because of the over-sampling the sample of farms exhibits the U-shape that characterizes the global relationship between agricultural productivity and scale. Population representative household surveys in low-income countries tend to contain few if any large farms. As indicated in Figure 1, there are few farms even above 10 acres in such environments. The U-shape is simply not visible in low-income country rural data sets because of survey design.⁹

There are measurement issues in most existing data sets as well that have made it difficult to identify the mechanisms that underlie scale economies that we focus on here. In many if not all low-income country data sets agricultural labor time is measured in days rather than hours. While time wages are generally paid on a daily basis for most agricultural operations, the true unit cost of labor time will be masked if there is variation in hours per day. The ICRISAT data record labor time use in hours and days. The data indicate not only that there is substantial variation in the average hours per day workers provide, but also that the amount of daily hours within an agricultural operation differs by operation scale.

Based both on the wage schedules provided by farm employers and the daily wages and

⁸Note, however, that weights are available so that the sample can be made representative of the population as well.

⁹To our knowledge, there are only two prior studies based on low-income country farm data that finds evidence of a U-shape. Kimhi (2006), using data on maize producers in Zambia, shows that diseconomies of scale characterize farms below 7.4 acres, which account for 84% of all farms, but that productivity rises with scale above that threshold. Muyanga and Jayne (2019), recognizing the representativeness sampling problem in existing data sets, obtain data from a dedicated survey of medium-sized farms and a representative sample of small farmers in Kenya that reside in the same villages. They also find the inverse relationship in the representative sample containing mostly small farms, but positive scale economies for the larger farms (25-124 acres), measuring productivity both as per-acre output and per-acre net returns, that is taking into account all input costs. However, neither of these studies provides evidence on the mechanisms behind the U-shape.

hours reported by farm workers, we show that the average hourly wage decreases with the number of hours worked, consistent with the existence of a daily fixed cost of employment. Based on these data, we are able to estimate the magnitude of the fixed components of daily wages paid by employers, which we show makes up over 50% of the daily wage paid to a full-time (8-hour) male wage worker. We also are able to document that smaller farms (but not the smallest) on average employ more low-hour hired labor across all of their operations than do larger farms. We show that as a consequence, the average hourly wage, inclusive of the imputed cost of family labor, increases and then decreases with farm scale. Consistent with this, we also find that for the same plot across time, when the amount of work increases due to more rainfall, on a smaller plot a higher average wage is paid while on a larger plot average wages are lowered.

Another major deficiency of existing data describing farming in low-income countries, even where mechanized equipment is used, is that there is little or no information on the capacity of farm equipment. The best surveys provide a detailed inventory of owned equipment by type (thresher, tractor, sprayer) and value, but little or no information on power or capability (e.g., horsepower, bushels processed per time unit). Thus, any empirical evidence on scale economies obtained from such data must assume that within machine types machinery capability is homogeneous - an eight-row harvester and a four-row harvester are not distinguished, even though their capabilities and suitabilities to different production scales are likely quite different. Most data sets also do not provide information on the time use and the rental price of equipment, by type or capacity. Thus it has not been possible to measure scale economies in farming due to economies of scale in machine capacity that could underlie the positively-sloped upper segment of the U.¹⁰

¹⁰There is evidence that larger farms and farms that become larger are more likely to be mechanized within countries (Zaibet and Dunn, 1998; Foster and Rosenzweig, 2011; Hornbeck and Naidu, 2014). However, in contexts in which all farms are mechanized, such as in developed countries, the mere use of mechanized equipment cannot by itself explain why larger farms are more productive than smaller farms.

The ICRISAT data too do not provide direct information on the power or capacities of the equipment that is used by the farmers. Tractors, for example, are not distinguished by horsepower or speed or towing ability. However, we show how it is possible to identify the varying capacities of one major type of equipment - sprayers - using the information provided on the amount of material sprayed and the time use of sprayers. This enables us to estimate an effective capacity function relating capacity- material sprayed per hour - to scale and to estimate the capacity pricing schedule.¹¹ We find that, consistent with sprayer capacity scale economies, larger farms do less spraying per acre and use higher-capacity and more expensive sprayers and we estimate that the implicit rental price of capacity declines as capacity increases. Based on our structural estimates we are able to identify the “optimal” scale of operation, conditional on existing wage rates and available sprayer technology, based on the sprayer scale economies - at 24 acres. This compares with the existing mean farm size in India of just over three acres, and is 2.4X the minimum optimal size derived, based on bullock technology, in the Zamindari Report. This estimated optimal scale is the scale at which additional increases in scale would lower productivity, but below which operational scale is too low and thus excessively labor-intensive, at least with respect to the control of weeds and insects.

Incorporating our structural estimates of sprayer technology and the fixed-cost specification of wage schedules, we calibrate the model by fitting its predictions to two moments of the data, stratified by area. Based on the calibrated parameters, we are able to reproduce the U-shape in profitability even when the production technology exhibits no scale economies. We are also able to identify the marginal product of labor in autarchic operations, which are not observed in the data. We show that, consistent with the under-utilization of labor in those operations, the marginal product in autarchic production is on average 40% higher than the hourly marginal wage rate. We

¹¹Our measure of sprayer capacity is identical to that employed by sellers of sprayers. We are thus able to compare our estimated capacity pricing schedule to those provided by sprayer vendors in the India and the United States.

are also to quantify the entry costs of a local laborer to the market, finding that is below the compensation paid by local employers, consistent with the hiring of non-local labor that is observed in the data.

Our counterfactual land consolidation simulation in which all farms are cultivating at the optimal scale and in which we allow for labor exit from the agricultural sector so that wage rates are endogenously determined indicates that there are 7.7 times too many farms in India. With all farms at optimal scale, output per acre is increased by 42% and output per worker by 68%. The principal sources of the gains are the elimination of labor misallocation due to the elimination of autarchic farming and the exploitation of machine scale economies. The new land distribution, resulting in an 87% decrease in farms, is also characterized by a reduction in the total labor force, but by only 16%. These results indicate that there is thus both an overall surplus of labor surplus and an underutilization of the existing labor force in agriculture.

In section 1, we describe the data and show that profits per acre exhibit a U-shape with respect to both farm size and plot scale that is robust to soil quality, crop choice, and farmer characteristics. We also present the evidence on and estimates of labor market transaction costs and non-linear machine pricing by capacity. In section 2, we set out the model, and in section 3 we carry out tests of the model based on the non-linear relationships between land size, production labor-intensity and average hourly labor costs. We also present the structural estimates of the sprayer capacity pricing schedule. Section 4 describes the calibration of the full model and presents the estimated parameters. In section 5 we assess the fit of the model to the data and discuss the estimates of the key unobservables, including the marginal products of labor on small, autarchic and large farms and the entry costs faced by workers in market participation. Section 5 contains the counterfactual land consolidation simulation based on the calibrated model and section 7 contains a summary of our findings and the potential implications of an endogenous consolidation of

landholding from the existing distribution that might arise if legal and institutional barriers to land transactions were eliminated.

1. The Data

a. Sampling and information content

We begin by describing the data we use. We do this for two reasons. The first is that we need to show the phenomenon that requires explanation, namely a non-constant relationship between farm productivity and scale that is not simply due to measurement error or omitted land quality. The second reason is that the data provide unique information on plot and farm location, input prices and input characteristics, which motivate (and permit) the new explorations of the underlying causes of non-linear agricultural scale economies.

Our principal data source is the six latest rounds of the India ICRISAT VLS panel survey, covering the agricultural years 2009-2014. The survey has two components - a census of all households in 18 villages in five states - Andhra Pradesh, Gujarat, Karnataka, Maharashtra, and Madhya Pradesh - and a panel survey of the households in those villages, which includes 819 farmers. A key advantage of the ICRISAT survey is that the sampling differs from almost all household surveys, which seek to achieve household representativeness, because the sampling frame is based on landholding size.¹² In particular, the survey contains in equal numbers landless households, small-farm households, medium-farm households, and large farm households. As a consequence of this sampling frame, we are able to examine both small and larger farms in a common environment, unlike in most surveys of farm households in countries with similar landholding distributions, in which most households own small plots.

The ICRISAT data are unique in other ways that are critical for identifying the underlying

¹²An exception is Myanga and Jayne (2019), which oversampled larger farms in Kenya and the ARIS-REDS surveys (Foster and Rosenzweig 2011), which oversampled large farms in 1967, and for which more recent follow up survey have been conducted.

mechanisms of scale economies. First, there is information on input quantities and prices by type of input, by farm operation and by individual plot collected approximately every three weeks.¹³ The high-frequency input information is thus likely to be more accurate than that found in almost all other surveys, which collect information once or at best twice in an agricultural season. Second, there is information on market input prices for workers, machinery, and animal traction collected at the village level, in addition to that elicited from the households survey, by work time. Third and importantly for identifying the role of mechanization in scale economies, there is information enabling the measurement of the power and capacities of machines. Fourth, there is information on how plots were acquired, e.g., inheritance or purchases, including information on the dates of inheritance, along with information on all other assets of the household.

b. *Descriptive information on scale and farm productivity.*

Figure 6 displays from the ICRISAT village Census and from the surveyed households in 2014 the cumulative distribution of farms by total owned (agricultural-use) landholdings along with the sample-household distribution of plot sizes. The figure shows that the full population (census) land distribution is similar to that of most low-income countries - 92% of land-owning households have less than 10 acres. Because of the sampling scheme, however, we observe detailed information on farms above 10 acres in the household sample - in contrast to the population distribution, households with more than 10 acres of landholdings constitute almost 40% of the sample.

The oversampling of larger farms is key to understanding the global relationship between farm productivity and farm scale. The sampling scheme provides the missing link between

¹³The size of the basic unit of operation, the plot, is not a choice variable - the size of a given named plot does not vary from year to year. Similarly, farm sizes are stable. There is little change in the number of plots owned by a farmer over the full span of the panel, from 2009-2014 - only 5.8% of plots were bought or sold, and the main reasons for any land turnover were inheritance or family transfer. Almost all plots therefore are inherited (0.74% of all plot observations involved a purchase of land). The means of land acquisition, including inheritance, does not differ by plot size. The 2014 Census data indicate the leasing market is only somewhat thicker than the land sales market, with 8.4% of landowners leasing out and 11.5% leasing in land.

developed-country large-scale farming and low-income country small-farm agriculture within the context of a single low-income country. This is because we are able to observe both the decline in profitability by scale, characteristic of low-income countries, and its rise with scale, characteristic of developed countries, in the same setting with comparable data across farms. Figure 7 displays the lowess-smoothed relationship between average real (1999 rupees) profits per acre in the main growing season (*Kharif*) and owned total landholdings for the full data set (2009-2014). As can be seen, as in most low-income countries, there is a monotonic decline in per-acre profitability with acreage below 10 acres. But then there is a monotonic increase, as is observed in developed countries.¹⁴

Using the detailed information of the data set, we can rule out three reasons for the U-shape that have been suggested in the literature, which has focused on the decline in productivity with scale below 10 acres: (a) measurement error in land size that is correlated with size, (b) land quality that is correlated with land size, and (c) credit constraints and farmer ability differences by farm size. Appendix A2 describes our tests of measurement-error bias, the influence of plot quality, and the roles of farmer ability and wealth. The estimates suggest that measurement error is small and is not systematically different by farm size, that plot quality, while an important determinant of profitability, is not significantly related to plot size, and that the U-shaped relationship holds across plots for the same farmer.¹⁵

Figure 8 displays three plots: the relationship between per-acre real profits and farm size repeated from Figure 7; that relationship estimated using the locally-weighted functional coefficient

¹⁴This is also true for output per acre as well.

¹⁵The ICRISAT survey farms are not especially fragmented. 43% have only one plot, and 74% have two or less plots, with larger farm having more plots (correlation = 0.6). The correlation between average plot size and total farm size is 0.7. We show below that the size of plots as well as farm size matter in determining scale economies.

model (LWFCM)¹⁶ from a specification including all of the soil characteristics in which the coefficients for farm size and the soil characteristics can vary non-parametrically with farm size; and the LWFCM-estimated relationship obtained solely from cross-plot variation within a farm and including the soil characteristics. All three plots display the U-shape, with the LWFCM-estimated curves obtained at mean soil and plot characteristics. Controls for soil quality evidently lower the observed profitability of the smallest farms (soil quality and size are negatively related among small farms) but have little effect on the upward slope. The within-farm estimates suggest that the U-shape also characterizes plots within farmers.¹⁷ Thus, variation in farmer wealth or farmer ability or heterogeneity in plot or soil characteristics do not explain the U-shape association between per-acre profitability and scale.¹⁸

The fact that the U-shape is seen in both plots and farms also suggests that the relationship is determined in large part by the size of the plot. This result raises two possibilities. One is that there are scale economies intrinsic to the production function itself and another is that there are scale economies in the sourcing of inputs for a particular plot. One way to distinguish these ideas is to examine whether operations on different plots are synchronized. If they are synchronized so that, say, a farmer can hire a worker to work on two different plots at the same day then the U-shape is likely to be an aspect of the production function. If they are not synchronized then the scale

¹⁶See Cai *et al.*, 2006. The specification we use is locally linear in profits and farm size.

¹⁷Interestingly, Assunção and Braido (2007) found, controlling for farmer fixed effects, an inverse relationship between output per acre and plot size but no uptick in productivity for larger plots using the initial ICRISAT data from the period 1975-1984. In that period the availability of mechanized implements was substantially lower than in the period covered by the latest rounds of the survey, 25 years later.

¹⁸The U-shape is also not due solely due to different crop choices by plot size. Using only plots devoted to cotton, one also observes the U-shape pattern of per-acre profits and plot size, as displayed in Appendix Figure A1. Cotton is the second largest cash crop in the ICRISAT sample, with 17% of all plots devoted to cotton. 20% of plots are devoted to soybeans, but soybeans are not grown on the very small plots that dominate the sample, so it is not possible to identify a U-shape relationship with that crop, but per-acre profitability on soybean plots rises with plot size above 10 acres.

economies may due to the different costs of applying inputs on plots of different size.

To gauge the synchronicity of plot operations, we exploited the daily calendar of operation start dates in the survey. We used these operations to compute the standard deviation of the start date of each operations across plots for farmers with two or more plots and the standard deviation of the distribution of average operation-specific start dates across farmers in the same village. Note that if in all operations the days of initiation were the same, the average standard deviations would be zero. Appendix Table A3 in the appendix reports these results, which show that the average standard deviations in operation start dates across plots for the same farmer are significantly different from zero and almost as large as those characterizing the synchronicity of operations across farmers.¹⁹

c. Autarchic farm operations.

Another feature of the ICRISAT farms exhibited in the data that is in common with farms in other low-income countries is the existence of operations in which no labor is hired and no family members work off the farm - the operation/farm is autarchic with respect to the labor market. The information in the ICRISAT data on monthly off-farm work by all family members and the information by operation on the employment of hired labor allows us to calculate a lower bound for the fraction of operations under autarchic production. We define an operation as autarchic if during the course of the operation no labor is hired and in the month the operation took place, no family member worked off the farm. This yields a lower-bound estimate of the number of operations that are autarchic because operation lengths are typically less than a month and we cannot know the labor-force status of family members for any specific period that is less than a month. It is possible

¹⁹ We posit two explanations for these patterns. First, farmers may find it desirable to diversify planting date in order to minimize overall risk and second, while pooling labor across plots may be advantageous when hiring labor, having asynchronous plots will be an advantage for the allocation of family labor on autarchic farms.

that in a given month some family member worked for wages but did not do so during a particular operation, so we miss some autarchic operations due to the mismatch between the periodicity of the off-farm family labor supply operation input information. Nevertheless, the lower bound based on these criteria is substantial - based on the sampling weights we estimate that 34% of operations were autarchic in 2014. And, autarchic plots are concentrated among the smaller but not the smallest plots, as shown in Figure 5, and was seen in the data on farms from Nigeria and China. A model of farm production needs to both account for the U-shape of per-acre productivity as well as the incidence of autarchic production.

d. *Fixed costs of labor hiring.*

In the next section we will set out a model to explain the U-shaped pattern of farm and plot efficiency as well as the existence of autarchic plots. We will focus on transaction costs in the labor market and scale economies in machinery capacity. We do this because these aspects of the farming environment are evident in the data. With respect to identifying a fixed cost component to hired labor, we use two sources of information. First, we use information from the price schedules for hired labor and labor plus bullock pairs by farm operation. Information on daily wages paid by farm operation was obtained monthly from one informant from each of the three classes of farmers in the 2010 and 2011 rounds of the survey according to the number of hours worked in the day. The reports are obtained at the beginning of each month over the full year. The second source of information is from the detailed information provided by each sampled farmer on wages paid and hours of hired labor, by type, for each operation and plot over the season.

The first salient feature of the data from the wage schedules is that a large fraction of workers paid daily wages work less than eight hours in a day. Figure 9 displays the distribution of hours worked in the day for hired male workers and, for comparison, bullock pairs and drivers, in the *Kharif* season for 2010 and 2011 combined. As can be seen, many workers are hired for less than

a full day - 31% of the daily wage reports for hired males were for workers who worked less than eight hours; for bullock pairs and driver, over 58% of daily wages paid were for work that was less than eight hours. This is in accord with the survey data on off-farm employment reported by respondents. In the 2014 round, for example, 44.4% of respondents working off farm for wages in agriculture operations during the peak *Kharif* season reported that their average working hours were less than eight.

The second feature of the data on wages and hours is that hourly wages differ by the amount of time worked. We computed hourly wages based on the monthly wage schedules and then regressed the log of the hourly wage for the two categories of hired inputs on whether or not the work done was for the full eight hours, with a full set of dummy variables for farm operation. Any hourly wage difference by daily hours hired could be due to low-wage operations occurring in slack periods with little work. The operation fixed effects ensure that this is not the case. The within-operation log wage estimates from the wage schedule data are reported in the first column of Table 1, where it can be seen that farmers pay an hourly premium for low-hour work - workers who work eight hours are paid a statistically-significant 33% less per hour than lower-hour workers; a hired bullock pair and driver working eight hours is paid over 22% less per hour than his part-time counterpart.

Data on wages paid and hours of work for hired workers as reported by farmers from the 2014 survey transaction files conform to these patterns. The survey data do not directly report how many workers were hired on a given task but they do report expenditure and hours for each task by demographic group (men, women, and children). We therefore focus on tasks in which there were less than or equal to 12 hours of hired male labor and assume that anything in the 8-12 range

represents full-time work by one worker and <8 refers to part-time work by one worker.²⁰ Because the equilibrium hours chosen by a farmer will itself be a function of the wage schedule, we instrument hours with the characteristics of the farmer’s plot, which determine labor demand but should not, net of hours hired, affect the wage in a competitive market.

The estimates in columns (2) and (4) in Table 1 are comparable to those from the monthly price schedules. In particular, there is 34.7% discount in the hourly wage rate for full time male hired workers and a 30% discount for full time bullock pairs with a driver. In the last column we use the same approach for one form of machinery, sprayers, for which we can control for an important sprayer quality - spray capacity, as discussed in detail below. The result, which controls for sprayer capacity, suggests that the scale economies in labor hiring do not also extend to machinery - there is no statistically significant difference in the hourly rental payment for sprayers by hours of use. This fact will be incorporated into our model.

The estimates in Table 1 are consistent with the existence of a fixed cost associated with hiring a worker for any amount of time and a fixed hourly wage. We can use the distribution of hours from Figure 9 and the wage estimates of Table 1 to construct an estimate of the fixed cost. Let the expenditure function for hiring one laborer for one day who works l_h hours be

$$(1) \quad w_h(l_h) = w_0 + w_1 l_h,$$

where w_0 is the fixed hiring cost and w_1 is the marginal hourly wage. Table 1 provides the wage “discount” for the average hourly wage in full-time work compared with that for part-time work (<8 hours). We need the distribution of hours worked by part/full time to compute the fixed costs because, as seen in (1), the fixed cost is a different share of the total wage for different levels of part-time work.. With f_i^p denoting the fraction of part-time workers working l_{hi}^p hours, the average wage

²⁰Note that if there are multiple part-time workers in the 8-12 category that would bias down an estimate of the part-time premium.

for part- time work given (1) is

$$\bar{w}^p = \sum_{i=1}^7 \left(\frac{w_0 + w_1 l_{hi}^p}{l_{hi}^p} \right) f_i^p$$

and, similarly, the average wage for full-time work is

$$\bar{w}^f = \sum_{i=8}^{12} \left(\frac{w_0 + w_1 l_{hi}^f}{l_{hi}^f} \right) f_i^f .$$

Table 1 tells us that $\bar{w}^p = 1.347\bar{w}^f$. Substituting, and using the marginal hourly wage rate of $w_1 = 21$ from the wage schedules to calculate w_0 ,²¹ yields an estimate of the fixed cost of 178 Rs per worker.

Is it plausible there are significant fixed cost components in labor costs, as implied by the labor price schedules and the transactions survey data? First, most agricultural laborers are hired on a daily basis. Farm operations are episodic and sequential, and operation timing is stochastic, depending on weather. Farm scale is too small for full-time work and contracting in advance is difficult given the vagaries of weather. So, each worker must be matched with a farmer who is seeking workers for a given day's task on a given day. Second, workers who do not have tasks on their own farm in a given day may be reluctant to take a part-time job on that day if a full-time job may subsequently be offered at the same hourly wage. Third, while this may not apply to very short jobs, there may be fixed costs such as the provision of an outside meal that must be incurred regardless of whether one works for, say, six or nine hours.

Finally, and perhaps most importantly, there are important transaction costs in the daily hiring of workers that arise from the fact that farms are spatially separated from where workers and farmers in the village reside; travel costs are thus not trivial.²² The ICRISAT' data provide the

²¹We calculate w_1 by determining the change in daily earning of an extra hour of work when working more than 8 hours.

²²Typically, farmers and workers reside in a village center with farm plots surrounding the residential area.

distance of each plot from the farmers' home (in the village center). The median distance is one kilometer.²³ If at least some of these turnover/search and travel costs are born by farmers, this will be manifested in hourly wage schedules that resemble those we see in the data. These fixed costs of hiring paid by farmers, moreover, may differ by land size, which will further affect the relationship between output and scale. First, empirically, in the survey data, the correlation between plot size and homestead distance is 0.36. Second, farmers cultivating larger plots may be willing to pay more up front to attract workers for a given operation, since the inability to hire would be more costly than it would be on smaller plots.

In Table 2, we report IV estimates with fixed effects for operation of the hourly wage discounts for full-day work differentiated by both plot size and plot distance for male hired workers and, again, for sprayers from the input transactions data. In the first column, we see that there is a statistically significant positive relationship between the wage discount and plot size. Inclusion of plot distance in the specification, as seen in the second column, does not eliminate the plot size gradient, but indicates that indeed, fixed costs of hiring rise with distance - for every kilometer the plot is located away from the homestead (village center), there is a 14% drop in the hourly wage associated with full-day work. The third and fourth columns of the table confirm that the rental price of sprayers, of given capacity, does not exhibit scale economies associated with hours of employment for any plots differentiated by size and/or distance.

²³The distance of plots to residences in the sample understates the average distance a worker must travel to get to an employer because a significant proportion of workers residing in a village work for a farmer located outside the village. The Yale EGC-CMF Tamil Nadu Panel Survey contains a representative sample of rural households in 200 villages in the Indian state of Tamil Nadu in 2011. In this sample, the median number of individual farmers that an agricultural laborer worked for in total during the main growing season (*kharif*) was seven. 23.6% of the survey respondents who worked for wages in agriculture reported working for a farmer located outside the village and 21.3% of farmers located in the villages who employed any agricultural laborers reported hiring laborers from outside the village. Among those workers traveling to a farm outside the village by foot or bicycle (63.8%), the average distance to the non-village farm was two kilometers. The median distance to a non-village farm for those traveling by bus (26.5%) was 8 kilometers.

e. Non-linear pricing of machinery capacity.

A second key component of our model is the existence of non-linear pricing in the cost schedules for farm equipment by machine capacity. The ICRISAT data, like most data sets, has only an incomplete description of farm equipment. The size or horsepower of tractors owned or used, for example, are not provided. The inventory of owned equipment in the 2011 round does provide, however, information on the price and horsepower of electric motors and submersible pumps. Figure 10 displays the relationship from the inventory data between the cost per horsepower and total horsepower for both sets of equipment. The figures clearly exhibit scale economies in pricing by machine power. The information on pricing, is insufficient, however, to demonstrate the importance of scale economies in equipment as an underlying mechanisms for farming scale economies, as we will show, because what is also required is the technical relationship between horsepower (in this case) and the ability of the equipment to accomplish agricultural tasks. We will demonstrate that there is sufficient information on sprayers, a key piece of farm equipment, from the input data files to pin down the relationships between machine power, machine capacity and price that is necessary for identifying the link between equipment and farming scale economies.

2. Model

a. CRS production and economies of scale in pricing.

In this section we set out the model that integrates both the fixed cost of labor and heterogeneous capacity in machines, based on the descriptive evidence for these two mechanisms. A central feature of both is that they induce scale economies not through the structure of the production, as is common in the literature, but via the costs of inputs. To highlight their role, therefore, we abstract from any scale economies in the production technology.

Although agricultural production takes place in stages, to fix ideas we focus on a one-stage agricultural production function with no scale economies. We assume that agricultural production is

described by a constant returns to scale production function g that consists of two inputs: land (a) and plant nutrients (e).

$$(2) \quad \psi g(a, e),$$

where ψ is a total factor productivity parameter, which we also assume is scale invariant. Thus, any scale economies in the model must come from input costs.²⁴

The amount of nutrients applied is itself described by a production process. For example, the application of fertilizer requires labor time. Removing weeds, which reduces competition for nutrients, can be accomplished using labor for pulling weeds, and/or by spraying, using labor and a sprayer.²⁵

As is consistent with the data on hourly wages by operation, we incorporate transaction costs associated with employment in the wage labor market and with hiring non-family workers. Workers entering the labor market for off-farm work on a given day face a fixed entry cost per day f as a result of transaction costs and/or travel (in effect we define a production stage as work done on a particular day). As a consequence, in equilibrium, farmers wishing to employ workers for just a few hours must at least partly compensate these workers for this fixed cost by paying a fixed fee w_0 .

The estimates in Table 1 pertain to the hiring of one worker for different hours of work. For a given operation, however, a farmer may need more than one worker, given a limit on the maximum hours any worker is willing to work. With labor compensation for one worker having a fixed and variable component, as depicted in equation (1), the cost of hiring l_h hours of work on a given day when each worker works only up to l_{mx} hours is

²⁴It is not clear, for example, why a plant may grow or more less rapidly based on the size of the plot on which it is located given the nutrients that the plant receives.

²⁵The nutrition interpretation of e is relevant for the analysis of scale in plant protection, which will be the focus of the first component of our analysis. However, in order to incorporate harvest work by workers and machinery in the analysis of overall worker demand we will subsequently expand the definition of e to include any form of inputs applied on the farm.

$$(3) \quad w_h(l_h) = \text{ceil}(l_h / l_{mx}) w_0 + w_1 l_h .$$

where $\text{ceil}()$ is the ceiling function. This structure says, for example, that a farmer must pay $w_0 + 4w_1$ for four hours of work but $2w_0 + 11w_1$ for eleven hours of work if 11 hours exceeds the maximum l_{mx} in a day that any worker will work. Equation (3) also characterizes the wage income for off-farm work by family members. Consequently, accounting for the entry cost, f , the opportunity cost of applying l_f units of family labor to the farm, is

$$(4) \quad w_f(l_f) = \text{floor}(l_f / l_{mx})(w_0 - f) + w_1 l_f$$

If $w_0 = f$, workers are fully compensated for the fixed costs of off-farm work.

Our production process allows labor and machinery to be substitutable in the production of nutrients. It also allows for the fact that machinery may be heterogeneous. Unlike for *manual* labor, where individual heterogeneity in productivity per unit of time (within gender) is relatively low and is in any case unrewarded in the market where time wages dominate (Foster and Rosenzweig, 1996), farm equipment devoted to specific tasks varies significantly in capacity and commands different prices associated with capacity.²⁶ Thus we need to distinguish machine time and machine capacity, with the farmer choosing both machine capacity, based on acreage and capacity prices, and how much time to employ the machine. We define capacity, consistent with definitions used for most farm equipment, as the amount of processed acreage a machine can accomplish per unit of time (e.g., acres covered per hour by irrigation or insecticide, acres of corn per hour harvested).

Thus we define the nutrient production function as

$$(5) \quad e(l, q, m) = (\omega l^\delta + ((1 - \frac{q}{\phi(a)}) q m)^\delta)^{1/\delta} ,$$

²⁶Farmers - that is the decision-makers - may differ importantly in capability relevant for making allocative decisions. We assume in the model, as is traditional, that all allocative decisions are correct, given technology and prices. We provide in an Appendix available from the authors tests of whether farmer ability is correlated with owned landholdings, exploiting information on inheritance of landholdings and rainfall histories. As noted, the within-farmer plot-specific relationships between profits and acreage indicate that any such a correlation is not solely responsible for the profitability patterns observed across landholdings of different size

where q is machine capacity and m is the number of units of time the machine is employed. The parameter ω captures the relative productivity of workers and machines, and δ captures the extent of substitutability between labor and machines.²⁷

Equation (5) embodies a relationship between machine capacity and *effective* machine capacity, which depends on farm size. We thus define the function $\phi(a)$ with $\phi'(a) > 0$ capturing the loss associated with using a large capacity machine on a small plot, so that effective capacity is $(1 - q/\phi(a))q$. For example, a sprayer that can cover a radius of ζ yards would be cost ineffective on farms where the radii of farmed area are significantly less than ζ yards, assuming that machine prices rise with capacity. Similarly, it is not cost effective to rent an 8-row harvester for land that has four rows of crops.

We also allow for economies of scale in farm machinery capacity arising from the cost structure of machine capacity. In particular, the rental cost per unit of time x_m for a machine increases at a decreasing rate with machine capacity:

$$(6) \quad x_m = p_m q^\nu,$$

if $0 < \nu < 1$ and assuming $q > 1$.²⁸ Finally, we assume that operation of the machinery requires θ units of family labor per hour of machine operation. Thus we subtract θm from the quantity of labor used to get the labor component of the nutrient production function.

Thus, profits are

$$(7) \quad \pi(a, l_h, l_f, q, m) = \psi g(a, e(l_h + l_f - \theta m, q, m)) - w_h(l_h) - w_f(l_f) - p_m q^\nu m.$$

Farmers maximize (6) subject to their family labor constraint that $l_f \leq l_T$, where l_T is the family

²⁷Some labor will be complementary with machine use, the labor used to actually run the machines, as we specify below.

²⁸Note that we have assumed there are no machine fixed hiring costs, consistent with the estimates in Tables 1 and 2. The existence of fixed costs in machine hiring would further increase machine scale economies.

labor endowment in hours.

b. Scale economies and labor-market transaction costs.

Before taking the model to the data it is helpful to understand the different ways that the cost structures for hiring labor and machinery contribute to scale economies. We begin with the labor-market entry costs. Setting aside machine use and thus machine scale economies and assuming $w_o = f$ ²⁹ profits are

$$(8) \quad \pi(a, l_h, l_f) = \psi g(a, l_h + l_f) - w(l_h) - w_1 l_f$$

The existence of the labor-market transaction costs gives rise to three regimes that depend on land size a differentiated by the allocations of family to farm and non-farm work and the use of on-farm hired labor.³⁰ In the first regime, at the lowest land sizes, family members work both on farm and off farm, as long as income from working both on and off-farm exceeds the income from on-farm work only. No workers are hired, given the transaction costs, and thus the upper-bound critical value a^* for this regime is where farmers do not hire workers and are just indifferent between entering the labor market and not.

In this regime and at the regime upper bound for a , the marginal value product of family labor is equal to w_1 :³¹

$$(9) \quad \psi g_e(a^*, e(l_T)) e_l(l_T) = w_1.$$

Thus, if all farms were in this small-scale regime, labor would then be allocated efficiently across farms.

²⁹When we calibrate the model, we allow fixed costs f to deviate from w_o .

³⁰A fourth regime in which farmers are both working off farm and hiring in workers on the same day would not arise when $w_o = f$. The farmer will replace hired work with additional family on-farm work that has the same opportunity cost at the margin as an additional family worker until hired work is zero, in which case the fixed cost of hiring a worker is also avoided.

³¹We omit the dependence of $e()$ on m and q as we are considering the labor-only case.

The second regime is where farm size is sufficiently large so that the profitability of employing all family labor on farm exceeds that from employing any family labor off farm but no hired labor is employed on farm. This is the autarchic regime, which arises because of the existence of the labor market entry cost that must be paid by the farmer when hiring labor. This hiring cost will make farmers reluctant to hire workers until land area reaches some higher threshold.

In the autarchic regime, starting at threshold land size a^* , given the fixity of family labor, the marginal product of labor *exceeds* the market wage and rises as land size increases. Profitability per acre falls with land size as the discrepancy between the (marginal) opportunity cost of labor w_1 and its marginal product rises - increasingly too little labor is used per acre as long as farm profitability without hired labor exceeds profitability employing farm labor and paying the added transaction costs. The fall in profitability per acre will continue until land size = a^{**} , where farmers are just indifferent between hiring a worker for sufficient hours l_h^{**} that the marginal product of labor equals the marginal wage, and defraying worker entry costs by working only with family labor. Assuming at most one outside worker is hired, a^{**} and l_h^{**} are defined by the pair of equations

$$(10) \quad \psi g_e(a^{**}, e(l_T + l_h^{**})) e_l(l_T + l_h^{**}) = w_1$$

$$(11) \quad \psi g(a^{**}, e(l_T)) = \psi g(a^{**}, e(l_T + l_h^{**})) - w_0 - w_1 l_h^{**}$$

The threshold land size at which any hired workers are employed is higher the larger is the transaction cost w_0 . Thus, the existence of transaction costs in the labor market is capable of explaining both the decline in profitability per acre with land size and the existence of autarchic farming at small, but not the smallest, land sizes - they both are manifestations of labor transaction costs. Note that if all farms were autarchic, agricultural labor would be underutilized and, given heterogeneity in family size and land size, labor inefficiently allocated across farms. In reality, the existence of at least some autarchic farms/operations means that the agricultural labor force is underutilized on average and mis-allocated across farms, with labor marginal products differing

based on land size and family labor force.

At a^{**} a worker is hired and the marginal product of labor falls to the market wage.

However, average labor costs rise at a^{**} because of the necessity of paying transaction costs, which are a large component of labor costs when hired workers are employed at low hours. Then, as land area increases above a^{**} , average labor costs fall, as the fixed component becomes a smaller share of labor costs, and profitability per acre rises as long as the hired worker is less than full time. If, conversely to our assumption in (3), the fixed cost were paid only on the first hired worker (labor hired in teams), then profits per acre would rise asymptotically reaching that for the smallest-acreage farms. Otherwise there would be new set of regimes associated with hiring the second and further workers. The existence of hiring costs can thus explain a partial upturn in profitability per acre. But even where one worker is part time so the marginal cost of additional work is w_1 the farm incurs the fixed cost for each worker; consequently, profits per acre must be strictly less than those that are achieved on the smallest farms in the absence of another form of scale economies.

c. Scale economies and machinery

Economies of scale in machinery can lead to additional scale economies by land size, which are governed by machine capacity and its pricing. Capacity is determined only by acreage, the parameter ν , and the ratio $\theta w_1 / p_m$.³² The overall productivity of nutrients and the substitution of labor and machines in the nutrient production function affect the hours of machinery use but not machine capacity. In particular, q , solves

$$(12) \quad \frac{(\phi(a) - 2q)\theta w_1}{p_m} + (1 - \nu)q^\nu \phi(a) - (2 - \nu)q^{\nu+1} = 0.$$

While a closed form solution for (12) is not generally available, it is evident that optimal capacity depends on the relationship between machine cost and capacity (ν) and on the effective capacity of

³²We assume labor utilization is infra-marginal rather than at the cusp where additional labor hours would require hiring an additional worker.

machines by area $\phi(a)$, as well as the different prices.

Scale economies in farm production associated with machinery thus require both that effective machine capacity depends on acreage and that there are economies of scale in machine capacity. If there were no cost advantage to using higher-capacity machines, even large farmers would use the smallest capacity machine. And if there were only a cost advantage to using larger-capacity machines but no relationship between effective capacity and area given actual capacity we would not observe small machines being employed on small plots. The cost advantage of using larger capacity comes from two sources in the context of our model: the rental price of machinery and the wage of machine operators. We assume that the rental price of a machine rises less than proportionately with capacity. We will obtain estimates of ν from actual price lists and from our survey data. These, indeed, indicate non-linear pricing of capacity. Second, we assume, based on the data, that farmers use one worker hour per machine hour ($\theta=1$) to operate the machine regardless of capacity.³³ Thus even if $\nu=1$ the combined hour labor and machine rental cost of a machine that has twice the capacity will be less than twice as much. The ν estimates will also allow us to identify non-parametrically the $\phi(a)$ function, as shown below.

The model implies that not only will the use of machinery increase with farm scale and with rising labor costs but so will the machine capacity chosen by the farmer. Implicitly differentiating (12), we get

$$\frac{dq}{da} = \frac{q^2 \phi'(a)}{(\phi(a) - q)(\phi(a)\nu - 2q\nu - \phi(a) + 4q)} > 0,$$

which is positive as long as $\phi'(a) > 0$ ³⁴, and

³³The survey data provides the number of hours of labor and the number of sprayer hours for each sprayer operation. In 2014, for all 850 sprayer operations average hours of labor use was 5.84 and that for the sprayer was 5.37.

³⁴This expression must be positive as the first-order condition for q implies $q < \phi(a)/2 < \phi(a)$ and the second-order condition requires $(\phi(a)\nu - 2q\nu - \phi(a) + 4q) > 0$.

$$\frac{dq}{dw} = \frac{\theta q(\phi(a) - 2q)^2}{p_m q^\nu v(\phi(a) - q)(\phi(a)v - 2qv - \phi(a) + 4q)} > 0.$$

The mechanism here is that at higher wages for the machine operator one wants to use fewer hours of machines per unit of nutrient added and this is only possible if the machine is higher capacity. On the other hand, an increase in the baseline cost of machinery tends to lower machinery capacity because it affects both the cost of capacity and the cost of machinery

$$\frac{dq}{dp_m} = -\frac{q(\phi(a) - 2q)^2 w\theta q^{-\nu}}{p_m^2 v(\phi(a) - q)(\phi(a)v - 2qv - \phi(a) + 4q)} < 0,$$

and a reduction in scale economies of will similarly lower capacity at any given scale of operation:

$$\frac{dq}{dv} = -\frac{q(\phi(a) - 2q)^2 w\theta q^{-\nu} \ln(q)}{v p_m (\phi(a) - q)(\phi(a)v - 2qv - \phi(a) + 4q)} - \frac{q(\phi(a) - 2q)}{v(\phi(a)v - 2qv - \phi(a) + 4q)} < 0.$$

As noted, the determination of machine capacity q is, in the context of the model, independent of the return to nutrients; it simply involves minimizing the cost of producing a unit of nutrients using a machine, given land area. On the other hand the optimal number of hours the machine of capacity q is employed depends on nutrient use and on the cost of labor used to provide nutrients not associated with machine use. For example, mechanized sprayers may be used to spread herbicide and thus control competition for nutrients by preventing weed growth. Alternatively, labor may be used to remove weeds once they have grown. The determination of optimal machine use is thus considerably more complicated than the choice of optimal capacity, and analytical derivatives cannot in general be signed.

3. Testing Model Implications

a. Unit labor costs and scale.

To test more directly that the U-shape in the marginal effect of land size on profits arises from changes in unit labor costs by land size, we first plotted the relationship between the real average hourly wage paid, which incorporate fixed costs, and farm size. In the ICRISAT data, family labor is priced at the marginal or eight-hour wage (as if fixed costs were fully born by the employer),

while hired labor is priced at the wage actually paid. Since the latter will be higher per-hour for low-hour hired labor according to the wage schedule, we should see that moving from the smallest farms to the largest, the average hourly wage first rises, as farms initially employ only family labor and then employ low-hour hired labor. At some threshold, the average wage paid falls as less low-hour labor is used. This is what we see in Figure 11.

To further test that the marginal land size effect on unit labor costs differs by land size, we estimated the relationships between the fraction of operations in the *kharif* season that employ low-hour ($\text{hours} \leq 6$) daily hired male labor, hired tractor services, and hired bullock pair services and the corresponding average hourly wages for each. We allow the marginal effect of land size to differ by land size by employing a quadratic in land. The estimates, which include village/year fixed effects and the plot characteristics, are reported in Table 3. For all three factors we see that the fraction of low-hour operations declines with farm size, and for both hired male labor and hired bullock pairs so do average wages across the range of farm sizes as captured by the quadratic specification. Thus, these estimates account for the rise in profitability per acre, and fall in unit-costs, above some threshold due to the declining use of high-cost hired labor and hired bullock pairs. The exception is for tractors, for which the effect is statistically insignificant but positive. This may reflect the fact that on larger farms more expensive tractors with more capacity are hired, an issue we will discuss below.

There are three limitations to the estimates in Table 3. First, there may be incomplete control for land characteristics, which may be correlated with land size and with input use. Second, the model and the labor cost figure suggests that a quadratic specification will not fully capture the change in the marginal effect of labor demand with farm size. Third, farm acreage is positively correlated with farmer wealth, so the acreage effects may in part reflect wealth effects. To remedy these limitations, we exploit the plot-specific panel feature of the data and intertemporal rainfall

variation to estimate, using plot fixed effects, the effects of rainfall on plot-specific input usage and average input costs by plot size. The plot fixed effects absorb any difference in plot quality and any differences in farmer permanent wealth, associated, for example, with the value of her total landholdings.

For most levels of rainfall in the semi-arid tropics in which the ICRISAT farmers are located, increases in rainfall increase input productivity and thus should increase input use. The exceptions are inputs that are employed in the planting stage, which principally occurs before the major component of the rainfall realization is known. Tractor use is mostly confined in the sample to planting-stage operations (tillage, plowing). Thus we use tractor employment as a placebo - rainfall should neither affect tractor hours nor the average per-hour rental price of tractors. On the other hand, for small plots higher levels of rainfall will increase average hourly wages if the additional rainfall induces the hiring of low-hour post-planting labor while for the larger plots, in the same rainfall area, increases in rainfall induces a shift from low- to normal-hour operations and average input costs decline.

We first establish that rainfall does indeed increase plot-level productivity and affects the demand for inputs. In the first column of Table 4 plot and year fixed-effects estimates of rainfall and rainfall squared on *kharij*-season profits from each plot are reported. As expected, increases in rainfall increase profits. And, in the second column, the estimates indicate that increases in rainfall also increase the number of hours of hired labor employment and hired bullock pairs, but the latter effect is only statistically significant at the .07 level (one-tailed test), consistent with bullocks being primarily used in the early stages of the production cycle. The effect of rainfall on tractor hours, as expected, is not statistically significant by conventional standards and is economically insignificant as well. In parallel, an increase in rainfall decreases the costs of both average hired male labor and bullock rentals, but has no effect on the hourly cost of tractors.

Having found that variation in rainfall on a given plot affects its profitability, the number of hired labor hours, and per-hour hired labor costs on average, we then estimated the effects of rainfall on the fraction of operations on the plot that employ low-hour hired male labor and the average wage paid by plot size, using LWFCM. The plot and year fixed-effect estimates of the effects of rainfall at mean rainfall by plot size on low-hour labor use, and the associated 95% confidence interval, are reported in Figure 12. The figure is consistent with the shifting of regimes of labor employment in the model - at small plot sizes, increases in rainfall statistically significantly increase hired low-hour labor use while for larger plots low-hour labor operations are statistically significantly reduced when rainfall increases.³⁵ And in Figure 13, among the larger plots increases in rainfall statistically significantly reduce average hourly hired labor costs.³⁶

b. *Identifying Equipment Scale Economies as a Source Farm Scale Economies: the Case of Sprayers.*

In this section we adduce indirect and direct evidence on machine scale economies, inclusive of estimation of capacity pricing schedules and the parameters of the effective capacity function, the fundamental components of such scale economies. As in most data sets describing farming where mechanized equipment is used, we see that larger farms are more likely to be using mechanized farm equipment, as shown in Figure 14 for tractors and sprayers. There has been scant evidence, however, on the relationship between machine capacity and scale.

The ICRISAT transaction data on usage of machinery, as seen in Figure 15, is suggestive of the rise in machine capacity with farm scale - while average hours of equipment use per acre first increases with scale, above 12 acres per-acre use of both types of equipment *declines* with farm size. This decline in machine use on a per-acre basis as farm size increases among larger farms is

³⁵Because, as noted, plot size and farm size, and thus farmer wealth, are positively correlated in the data, the rainfall coefficients at small plot sizes may be underestimated due to credit-market or liquidity constraints on the ability of small farmers to employ additional hired labor. More relaxed liquidity or credit constraints for larger farmers, however, cannot explain the negative effect of rainfall on per-unit labor costs.

³⁶The effects of rainfall on per-acre profits does not vary by plot size over the full range of plot sizes.

consistent with machine capacity scale economies. However, these patterns are not directly informative about whether machine capacity actually increases with farm size, whether there are scale economies in capacity due to non-linear capacity pricing, or at what scale, if any, capacity scale economies dissipate completely. To address these issues we need a measure of machine capacity.

The capacities of machines used by farmers are rarely, if ever, available in data sets based on household surveys from low-income countries. The ICRISAT data set is no exception. However, the detailed information on materials purchased and on hours of machine use by machine type in the ICRISAT transaction modules permits the computation of capacity for one type of equipment - sprayers. This is because there is information on the amount of material sprayed - weedicide and insecticide - as well as information on hours of sprayer usage by plot and operation. These data can thus be used to compute capacity - amount sprayed per hour. Sprayer capacity is typically given in spray rates for a given nozzle size - material volume per time unit. The relevance of this measure for farming scale is that flow rates translate directly into area sprayed per hour, given a target amount of material per area.

Another advantage of sprayer technology is that we can exploit the information on input use by operation to directly measure the labor savings from spraying. This is because an important alternative to spraying for protecting plant nutrients is weeding, which is typically done manually. The data provide the hourly rental rate for the sprayer used and labor usage for both spraying and weeding operations. Figure 16 displays the relationships between per-acre expenditures on sprayers and on the labor used in spraying and weeding. As can be seen, as farm size increases farmers are using more expensive, and thus, as we will see, higher-capacity sprayers;³⁷ and weeding labor costs per-acre plummet while per-acre labor time used in spraying only slightly increases with scale.

³⁷The per-hour price of the sprayer used increases monotonically with acreage. As noted, the average hourly wage falls with acreage at medium farm sizes and is constant at larger farm sizes.

We focus on sprayer technology so that we can for the first time obtain direct evidence from data in a low-income setting on how capacity heterogeneity in equipment contributes to economies of scale in agriculture that persists above a farm (plot) size threshold, conditional on machine use. This, of course, does not mean that sprayer technology is the sole source of economies of scale due to mechanization. But spraying weedicide and insecticide is an important operation. Spraying labor costs alone account for 13.6% of total input costs in the *Kharif* season. And as can be seen in Figure 15 there are more hours of use of sprayers at every land size than hours of use for tractors, the next most used machinery.

The data on asset ownership from the last round of the ICRISAT survey indicates that ICRISAT farmers use two types of sprayers - manual sprayers, whose median cost in 2014 is 700 rupees, and power sprayers, whose median cost is 2700 rupees. Of the 10% of farmers who own a sprayer, 25% own a power sprayer. Table 5 reports, based on these data, regressions of tractor ownership, ownership of any sprayer, and, conditional on owning a sprayer, ownership of a power sprayer on land size and land wealth (based on the farmer's own assessment of the rental value of the land). As can be seen scale, not just wealth, matters for equipment ownership: net of wealth, farmers with more land area are more likely to own a tractor and sprayer; and they are more likely to own a power sprayer if they own any sprayer.³⁸

Even among power sprayers, however, there is heterogeneity in capacity. Table 6 provides information taken from the web site of an Indian purveyor of power sprayers (KrisanKraft) that provides power sprayer prices by precisely the measure of capacity we can construct from the ICRISAT data - the amount a sprayer can broadcast in litres per hour. A notable feature of the

³⁸The market value of land fully captures land quality in a well-functioning market. However, the effect of scale conditional on the value of land may under-estimate scale economies because the value of land may reflect in part the existence of scale economies, with per-acre rental prices varying positively with land size. Over all acreages we find that the reported rental prices of individual plots do indeed rise with total acreage.

listing is that there are substantial differences in sprayer capacities - the spray rate of the highest-capacity power sprayer is over 13 times that of the lowest-cost model. More importantly, the posted price schedule exhibits economies of scale in sprayer capacity, the sprayer price per unit of capacity decreases as capacity increases.³⁹

The survey data on input usage suggest that farmers are exploiting economies of scale in spraying. Table 7 reports village/year fixed effects estimates of the effects of land size on the use of any sprayer, on weeding hours per acre, on sprayer hours per acre, on the log of the price of the sprayer used, and on the sprayer flow rate (capacity) net of the effects of the land quality variables. These estimates indicate that net of year-village effects, larger landowners use pricier and higher-capacity sprayers and larger landowners spend less time per acre in *both* spraying and weeding compared with smaller farmers.

To test directly for scale economies in spraying and the limits, if any, to sprayer scale economies we use the structure of the model, simultaneously estimating the effective capacity function $\phi(a)$ and the key parameter of the price function ν from the information on the capacity and per-hour rental prices of the sprayers used by the ICRISAT farmers. The challenge for estimation is that we observe only the capacity of the sprayer that is used by the farmer: capacity and the per-capacity price are choice variables. A positive shock to equipment prices for example will lower a farmer's selection of machine capacity, leading to a negative bias in the estimated parameters of the cost function. As a consequence we employ GMM using land area and land area squared as instruments.

Equation (11) solves for the optimal choice of q and embeds within it the effective capacity function and the capacity pricing parameter ν . It also contains, however an additive term that

³⁹We will formally test below whether the sprayer capacity ν parameter is less than one in the KrisanKraft price schedule and in the ICRISAT data based on the sprayers rented by the farmers.

includes the village wage rate w and the base price p of the capacity pricing schedule, which may be endogenously determined. We thus rearrange (11) and difference across randomly selected pairs of households i and i' in each village j to eliminate w_1 and p_m . In combination with the differenced log of (5), we then have moment conditions of the following form

$$(13) \quad E \left(\frac{(1-\nu)q_{ij}^\nu \phi(a_{ij}) - (2-\nu)q_{ij}^{\nu+1}}{(\phi(a_{ij}) - 2q_{ij})} - \frac{(1-\nu)q_{i'j}^\nu \phi(a_{i'j}) - (2-\nu)q_{i'j}^{\nu+1}}{(\phi(a_{i'j}) - 2q_{i'j})} \middle| a_{ij}, a_{i'j} \right) = 0 \quad .$$

$$(14) \quad E \left(\ln(x_{ij}) - \nu \ln(q_{ij}) - \ln(x_{i'j}) - \nu \ln(q_{i'j}) \middle| a_{ij}, a_{i'j} \right) = 0 \quad .$$

We parameterize $\phi(a) = b_0 + b_1 a + b_2 a^2$ and employ GMM using land area and land area squared as instruments to and the b_k . The errors over which these expectations are taken would be variation in the effective price of the machinery, say due to demand at the time one is undertaking a task, which we are assumed to be uncorrelated with differences in farm area within a village.

If $b_1 > 0$ and $b_2 < 0$, we can identify a maximum farm scale at which further increases in acreage could not exploit *existing* equipment scale economies. This establishes the land size at which per-acre profits are maximized - the technologically-determined optimal land size at which further increases in scale would not increase productivity but below which, given that $b_1 > 0$, farms are less productive net of costs. It is important to note that we can only identify the farm scale upper bound, if any, based on machinery that is actually available to the ICRISAT farmers. Machinery scale economies for farm sizes beyond the maximum scale of the farms in the population, if any, cannot be estimated because machinery for such farm sizes would not be marketed - they would not be available for rent or purchase. Farmers (and policy-makers) would thus have limited knowledge of how expansion of scale beyond that in the population would reap benefits via machine scale economies.

Table 8 reports the GMM estimates of the effective capacity and pricing function parameters

and their robust standard errors. All parameters are precisely estimated. We can reject the hypothesis that $\nu = 1$ and thus that there are no scale economies arising from the cost of higher capacity machines. We can also compare our estimate of ν based on the sprayers used by the ICRISAT farmers to that characterizing the price schedule for the four power sprayers sold by KrisanKraft, listed in Table 6, and to that for four power sprayers offered in the United States, as surveyed in Stiles and Stark (2016). The estimated ν 's are reported in Table 9. As can be seen, our GMM estimate of ν for the sprayers used by ICRISAT farmers is comparable to that for the KrisanKraft sprayers that are sold across India. However the estimated India sprayer ν is more than double that for the sprayers sold in the United States - economies of scale are evidently substantially greater for the power sprayers available in the United States.

The estimates of the b_k indicate that $\phi'(a) > 0$ over the relevant range. Thus, smaller farms are less cost-effective than larger farms, given that ν is substantially less than one. The estimates also indicate that there is a land scale at which effective capacity reaches a maximum. The point estimate of the maximum is 24.5 acres, with a 95% confidence interval of ± 3.6 acres. The 2014 Census of all households in the 20 ICRISAT VLS villages indicates that only 1.1% of households owning land have total landholdings above even the estimated lower bound of the maximum (20.9 acres). As expected, there are essentially no farms that could exploit further scale economies, given the sprayers that are available. The estimated peak maximum land size does not suggest, as noted, that larger farms than are observed in the ICRISAT area would not be more productive; rather it is consistent with an equilibrium in which none of the largest farmers has an incentive to expand given the available sprayers in India. Of course, that most farms are below this maximum, conditional on the local availability of machinery, implies that there are other barriers to land consolidation, resulting in an excess number of farmers. We will discuss these in the conclusion.

4. Model Calibration

In this section, we calibrate the full model, using our estimates of the sprayer pricing functions and other moments from the ICRISAT data. There are three principal aims: first, to ascertain if our model structure using empirically-validated and plausible parameters reproduces the observed U-shape in per-acre profitability and other moments of the data; second, to calculate the optimal size of farms using existing machinery technology and pricing in India; and third, to carry out a counterfactual in which we move from the existing distribution of farm sizes to one in which all farms are at the estimated optimum size, conditional on existing machine technology in India.

Calibration of the model also enables us to estimate separately hiring fixed costs w_0 and the cost of labor-market entry f , because they are identified from multiple moments of the data, and to estimate the marginal product of labor under autarchy, in which the family labor marginal product, which exceeds the marginal market wage, is not directly observed.

As noted, to fix ideas we used a one-operation, one day farm model and focused on sprayers to identify machine scale economies because it is one of the few cases in which machine capacity can be measured explicitly. However, in order to realistically calibrate the model to the relationships between land scale, profits, output and labor that describe multiple operations with a variety of machinery we need to specify how many days of work there are for all operations and how the demand for farm labor, at the aggregate level, translates into the demand for workers at different scales and prices of labor.

We proceed as follows. First, we augment the model to allow for multiple days of inputs. Second, because we do not have information on machine capacity for all machinery we assume that the basic structure of scale economies in machinery and employment are the same for each operation and that the contribution of each operation to overall production is similar. As a result, the model can be collapsed into a single decision that is then replicated on each day in which the farmer works on his farm. The model will thus depict a representative operation. We work with the

plot level data to identify the number of such operations and days per operation over the course of the season by plot size. The numerical model is solved such that, for given parameters and plot area, the farmer's optimal allocation of inputs is determined. The model is nested in a method of moments estimator that relates key features of the data to the predictions from the farmer's problem. The target moments are the profitability per acre and output per worker by total acreage. We assess the fit of the model to a series of non-targeted moments, including labor and machine utilization by farm size.

The particulars are as follows. First, the multitask production function is assumed to be

$$(15) \quad \psi \tilde{g}(a, e_1 \dots e_D) = \psi a^\alpha \prod_{i=1}^D e_i^{(1-\alpha)/D}$$

where D is the total number of days worked and e_i denotes nutrients provided on day i .⁴⁰ Plant nutrients e_i are produced according to (5) and face the same price every period, therefore $e_i = e_j \forall i, j$ so in the context of the model (15) collapses to

$$(16) \quad \psi \tilde{g}(a, e_1 \dots e_D) = \psi a^\alpha e_1^{1-\alpha}.$$

To compute the number of operations we take advantage of the fact that the VLS data provide plot-level inputs by operation and start date. On average we find that there are 12.75 operations per plot with the inter-quartile range running from 9 to 16. Unfortunately, the operations data do not indicate how many days were used for each operation. However, we can identify the days of work on a farm by family workers. Restricting attention to households with one male worker and assuming that the family worker will be working full time whenever hired workers are employed we estimate that the number of days per operation is one for the smallest farms, rises to about 1.5 at 5 acres and then levels out. On this basis, we took 21 days as a reasonable estimate of D .

⁴⁰In practice, of course, this implies that the elasticity of substitution with respect to plant nutrients on different days is one. In practice, of course, days in the same operation would be more substitutable and days in different operations more complementary. However, given evidence on days per operation presented below this seems a minor assumption that has a big effect on tractability.

We use the estimated coefficients for $\phi(a)$ and ν and the estimate of the marginal wage based on the wage and hours data of $w_1 = 21$. We establish from the data that on average we have three potential agricultural workers per household who work on average 8 hours per day so we set full time work hours l_{mx} to 8 and the daily family labor endowment l_T to $3 \times 8 = 24$ hours. The vector of remaining parameters that are calibrated is $[\alpha, \psi, w_0, f, \delta, \omega, p_m, n_0]$.

We make two additional adjustments that take into account the existence of multiple operations that may hire multiple workers. First, we allow for the labor transaction cost to vary linearly with farm acreage. While the estimates in Table 2 suggest that these costs per hired worker rise with land size, it is possible, for example, that are scale economies in the recruitment of transportation of multiple workers or if workers hire themselves out in teams. We thus do not constrain this coefficient to be positive or negative.

Second, when there are multiple family or hired workers the respective cost functions will have jumps associated with the employment of full time workers. Just as a farmer will avoid hiring a first worker as acreage expands until the productivity loss from not doing so exceeds the fixed cost, he will avoid hiring a second worker until the productivity loss of having only one worker exceeds the cost for hiring the second worker. This complicates numerical optimization by creating multiple local maxima. It also seems unrealistic in that we are focusing on a representative stage that captures multiple different stages with different daily labor needs. We thus smooth the hired and family labor cost function with the lower envelope of the exact equations specified above:

$$(17) \quad w(l_h) = (w_0 + n_0 a)(\mathbf{I}(l_h > 0) + l_h / l_{mx}) + w_1 l_h .$$

$$(18) \quad w(l_f, a) = (w_0 + n_0 a - f)l_f / l_{mx} + w_1 l_h$$

where n_0 is the effect of acreage on the fixed cost paid to workers and, as noted above for family workers, we have set $l_{mx} = 8$ as the maximum workers will generally offer to work (the mode of

the hours distribution). Note that the jumps in the exact model only apply to the last worker hired but we are fitting to profits per *acre* and output per *worker*. Therefore, as acreage grows, the difference between the exact and smoothed models in profits per acre and output per worker gets smaller and smaller.⁴¹

We solve the model based on the assumption that farmers maximize profits conditional on farm area and family size minimizing over the parameter estimates the sum of squared distance between the predicted values for profits per acre and output per worker and those values in the data at twenty different levels of area.

Although the seven parameters are being calibrated jointly it is possible to get some sense of the identification of specific parameters. We assume, as is apparent in the data, that machine use and hired labor are minimal over the regime in which family members are working off-farm and incorporate the smoothed return to off-farm work. Then, given the assumed production function, l , and w_1 , the parameters ψ , p_m , α_i , $w_0 - f$, n_0 can be identified from profits per unit area and output per worker on very small farms and those just at the cusp of autarchy (see Appendix A2). The size of the fixed cost is then determined, given these parameters, and the minimum of the profits-per-acre curve, which corresponds to the point at which the farm first hires workers. The elasticity of substitution is identified from how the changes in the productivity of machines (as dictated by the capacity schedule) translate into changes in demand for workers as acreage increases.

⁴¹Numerical calculations using the estimated parameters indicate that the discrepancy in profits per acre at 24 acres, for example, is 5.8 Rs. And while there is a difference in the optimal number of workers reflecting the tendency to hire only full time workers in the non-smoothed case, the implications the shapes of the optimal profits per acre and output per worker curves is essentially the same. The discrepancies are higher among farms where the family is supplying off-farm labor. However, it is important to note that profits per acre, one of our targeted moments, is calculated by assuming a fixed opportunity wage so the profit data would not fully capture the implied jumps even if the exact wage schedule were fully correct for on-farm work. Moreover, the exact schedule implies that many off-farm workers would want to work just an hour in the labor market (to benefit from the fixed cost component). But such short employments are not visible in the data nor implied by the model of hiring. We thus see the smoothed wage function as a reasonable as well as numerically convenient approximation at the low end.

The price of machinery and relative worker productivity come more broadly from the overall pattern of profits and output per worker on larger farms, where machinery is used, relative to smaller farms, where it is not.

The calibrated structural parameters values are

$$[\alpha = .381, \psi = 484, \omega = .880, f = 65.5, \delta = .934, p_m = 20.8, w_0 = 156, n_0 = .761] \quad .$$

Of primary substantive interest are the fixed cost of hiring workers, the incidence of these costs and the substitutability of labor and machines. The latter is quite high. The estimated coefficient on δ is consistent with an elasticity of substitution of 13.9. The fixed cost paid by farmers for a hired laborer w_0 is 159 for the mean-size three-acre farm, consistent with our calculation of w_0 based on the regression estimates across all farms and the distribution of observed hours for a given worker. And, consistent with the wage and hours estimates for hiring a single worker from the transaction data in Table 2, w_0 is increasing in acreage. It is worth emphasizing that the sources of variation used for identifying these two estimates are quite different: there was nothing, for example, in the matched moments from the structural analysis that distinguishes full and part-time work or family and hired workers. Interestingly, the estimated fixed cost f incurred by workers in the village who work in the village of Rs 41.7 (or just under two hours at the variable component of the wage), which cannot be observed in the data, is less than the fixed cost component of compensation. Because the fixed cost paid through the wage will reflect the opportunity cost of the last worker hired to meet demand and this worker may be from outside the village, local workers who will have to travel less far may be more than compensated for their lower costs of being employed in the village.

5. Model Fit and Implications

Figures 17 and 18 present the fit of the data and model to the two targeted series. In Figure 17, profits per acre, the model captures both the initial decline and the subsequent rise in profits per

acre, with both constant technical scale economies and scale-invariant TFP (ψ). As might be expected given the lack of heterogeneity in the model the data changes a bit more sharply than does the model. Variance in, for example, labor productivity in different stages would cause the minimum in the model to move to higher or lower acreage. But the minimum in the model for our representative task corresponds well to the middle of the lower plateau in the data. Similarly, the model does not capture the extreme right of the data. For larger farms, the pattern predicted by the model is governed by the quadratic capacity curve $\phi(a)$ that is estimated from the sprayer data. In practice, one might expect the curve to reach a maxima and then flatten out. But we cannot test this idea given that we have so few farms of this size or greater.

Output per worker (Figure 18) also fits well. There is a strong upward trend predicted by the model as well as in the data, with the largest farms exhibiting output per worker that is almost twice that of the smaller farms. There is also a distinctive peak in the model at around 9 acres, which would be just below the point at which the farmer first decides to hire workers. While, as with profits per acre, the maximum output per worker in the data is much smoother, it rises, relative to trend, at essentially the same point as the one predicted by the model. Again the fit seems to diverge most among the largest farms after the capacity curve reaches its maximum.

We did not directly target the total number of worker hours. While this measure is implied by profits per acre and output per worker when no machinery is employed such as in the lowest farm sizes, the calculation of hours at the upper end depends importantly on the structure of the model and implications on machine scale derived from the sprayer data. Figure 19 shows the resulting fit, which also looks quite good. The overall shape is similar and, while the overall estimates from the model are systematically lower the difference is only about 10 percent for both small(1 acre) and large (24 acres).⁴²

⁴²The model does less well in predicting labor per acre for very small farms. Our smoothing assumption implies that the marginal product of labor is constant (or slightly rising given that the fixed costs

We have already noted that the calibrated model estimate of the fixed component of the wage paid by farmers to workers matches well with that estimated from the wage schedules and that the unobserved fixed cost facing workers is actually below on average the fixed compensation component. The model also can be used to predict the fraction of farmers in autarchy in the labor market, something that, as noted above, is a common feature of farms in the developing world and that can only be approximated in the data. The model predicts that 33.8% of farms are autarchic given the distribution of farm size in the ICRISAT data, which is comparable to the lower-bound estimate of 34% of all operations, population-weighted, based on direct calculations from the survey data using input and labor supply data.

We can also use our calibrated parameters to estimate the marginal product of an extra hour of labor by farm size. The former is not observed in the data for autarchic farms that do not hire labor and in which family members do not work off farm. The marginal opportunity cost of employing an extra hour of a hired laborer also is not directly observed on all the other farms, as such costs must take into labor fixed costs. Estimates of labor marginal products by farm size are important as variation in marginal products is indicative of inefficiency in the labor allocation across farms. If all operations were of the same scale, for example, the marginal product of labor would be the same on all farms.

The marginal product of the *inframarginal* hour is the same for the smallest and largest farms and corresponds to the marginal hourly wage $w_1 = 21$ Rs, as observed in the data. The marginal product of an additional hour of work on autarchic farms depends, however, on the number of family workers, on acreage, and on all of the parameters describing the farm technology. For the

rises with acreage) over the range in which the family works. As noted this assumption corresponds to how costs for family work are costed out in the computation of profits. To capture this pattern the model would have to be constructed such that labor-market opportunity cost of on-farm work for small farms (where one may need just thirty to four hours over the whole season) is close to zero, as might be the case if work is done largely outside of normal working hours and thus has a low opportunity cost.

average family size of three workers, the calibrated parameters estimates indicate that the maximum marginal product of workers on the over third of farms that are autarchic is 53.8 Rs., which characterizes that on 9-acre farms – the farm size threshold where hiring begins. Weighting by the actual distribution of autarchic farms by size, the estimated mean marginal product among autarchic farms is 43.9 Rs., more than double the marginal product of an infra-marginal hour on all other farms engaged as sellers or buyer in the labor market. These figures indicate that, given the large fraction of farms/operations that are autarchic, agricultural labor is not only mis-allocated across farms but is significantly underutilized.

As noted, among farms employing multiple family workers but no hired workers, the net compensation for the costs of wage labor market entry also can matter at some margins. And for the largest farms who can benefit from hiring more than one worker in an operation, fixed costs also matter. Using the smoothed estimates of the cost of labor incorporating the estimates of fixed and variable wage costs that assume a uniform distribution of daily hours for hired workers, we find that among farms with family members working off-farm the per-hour average marginal product of adding one worker per day is Rs 30.7, still 30% below the average marginal product of labor on autarchic farms. This figure corresponds to the variable component of the wage w_1 plus the hourly component of the difference between the fixed cost of working off farm and the payment made to the worker by the employer. For the largest farms (24 acres) the per-hour marginal product of adding one worker per day is Rs. 43.1 on average, reflecting only the fixed wage component w_0 and our empirical finding that the fixed component of wage payments is higher on the larger farms.

The model also allows us to determine how per-acre profitability, output per worker and employment per acre would vary by land size in the absence of machinery, but retaining labor-market transaction costs. Figure 17, which shows how profits per acre evolve with land size, indicates that the entire upswing in profits per area for large acreages is due to machinery. The lack

of an increase in per-acre profitability by scale without machines among the larger farms is the result of the fact that on these farms we found that labor fixed costs rise with acreage - there does not appear to be team hiring to avoid worker transaction costs.

Figure 18 shows that machines play an important role in expanding output per worker. Output per worker in the presence of machines increases with acreage, only leveling off at the point where capacity is maximized. But output per worker without machines is essentially flat after the farm enters the regime in which it is hiring workers. Either in terms of profitability per acre or output per worker there is little benefit to expanding acreage per farm if one cannot take advantage of the scale economies inherent in machinery.

Finally, Figure 19 illustrates the point that workers and machines are important substitutes. Absent machinery, the number of workers employed per acre is about 50% higher than it is when farmers can use available machine technology in the regime in which workers are hired. For very small farms, where no machinery is employed even when machines are available the two curves coincide, as is the case for profits per acre and output per worker.

6. Redistributing Land So That All Farms Are Conditionally Optimal-Sized

In this section we use the calibrated model to carry out a counter-factual in which we shift the existing distribution of farms to one in which all farms are at the optimal size conditional on the existing machine technology in India. By shifting to a uniform farm size distribution we eliminate the mis-allocation of labor across farms due to labor market transaction costs. Shifting to a uniform farm size that also fully exploits the existing scale economies in machinery maximizes the return on land.

As is evident from the labor hours per acre graph, an expansion in land size reduces workers per area at the given current wage. It is not likely, however, that the rural wage would stay the same if there were a substantial reduction in farm labor utilization. It is thus necessary to take into account

how a reduction in worker hours would affect the equilibrium wage. To calculate the optimum farm size conditional on the existing machine technology of India we thus embed the model in an equilibrium framework. This is because optimum farm size depends on the market wage rate, which will depend on the total number of workers that stay in the agricultural sector given the changes in agricultural worker demand. Based on data describing the existing work force size and total hours of employment, in combination with an estimate of the urban labor demand elasticity, we can estimate, using the calibrated parameters of the model, the wage rate, profits per acre, and output per worker corresponding to any change in the distribution of land sizes in agriculture.

There are three steps. First, for any existing wage we calculate the farm size that would be optimal. We do this by calculating the acreage using the model parameters that maximizes for any given wage the profitability per acre. Second, we translate the reduction in hours associated with the change in farm size into reduction in workers over the season. If farmers used the same number of workers every day then this would be straightforward. In practice, as we saw, farms undertake specific operations and it appears that at least for plots of the size seen in the data each operation takes one to two days. To make this adjustment we turned again to the ICRISAT sample and constructed population estimates of total hours and of the size of the agricultural work force using the sample weights. We obtained 1023 agricultural workers (cultivators plus agricultural wage workers) and 188,101 farm hours as characterizing the current distribution of land. Assuming the ratio of workers to worker hours is independent of the distribution of farm sizes, because under existing technology the number of agricultural operations is independent of scale, we can use the projected total hours from our model to compute the number of agricultural workers that must be available over the agricultural season for any given land-size distribution.⁴³

⁴³This calculation misses people who are working in a village but live elsewhere and the workers who reside in the village but work outside the village but if these villages are representative of neighboring villages this difference should be a wash.

A key assumption about any redistribution of land is how the resulting change in the demand for agricultural labor affects the equilibrium wage. In Arthur Lewis' (1954) pioneering work on surplus labor he assumed that the urban sector could absorb as many workers as would exit from agriculture with no wage decline because of the absence of a fixed factor in the urban sector. However, if the urban or industrial demand elasticity is not infinite, we need to calculate the decline in the rural wage that would be necessary to induce these "extra" workers to exit from the agricultural sector so as to equate supply and demand in the agricultural sector.

To compute the equilibrium conditional "optimal" land distribution that takes into account general-equilibrium wage effects under a regime in which there is an elastic urban labor demand curve we used the estimate from the literature of -0.4 and assumed that the rural wage decline to maintain equilibrium would be proportionate.⁴⁴ We obtained from the 2011 India census an estimate of the number of farms in India and the number of urban workers. This then allows us to translate our estimates from the ICRISAT data into national statistics so that we can calculate the growth in urban workers due to the reduction in agricultural worker demand. Combining the wage elasticity estimate with the size of the urban work force we can then determine a fixed point where the rural wage and urban wage change proportionately⁴⁵ for any given distribution of land.

The first column of Table 13 reports the existing characteristics of Indian agriculture, including the mean farm size, number of farms, and the size of the agricultural work force based on the 2011 Census data and our calibrated estimates for output, per-acre profits, and machine use. The second and third columns report, respectively, the Lewis surplus-labor equilibrium and the

⁴⁴Table B2 in Lichter *et al.* (2015) contains a meta-analysis regression coefficient for India of -0.398. In Goldar (2009), the most recent estimates (1996-2003) for the short-run and long-run elasticities are -0.309 and -0.480, respectively.

⁴⁵We could assume that rural and urban wages must be equal before and after, but because our measure of urban labor demand is in the form of an elasticity, it is sufficient to assume that the ratio of the rural to urban wage stays constant allowing, for example, for persistent differences in the cost of living.

equilibrium in which there is urban wage decline that results from the shift to a farm distribution in which all farms are at the calibrated optimal size. The key difference between the two counterfactual equilibria is that, by construction, there is no decline in the agricultural wage rate in the Lewis equilibrium, while in the equilibrium we believe is more consistent with estimates of urban labor demand, there is a 8.6% decline in the wage. However, optimal farm size (24 acres) and the reduction in number of farms, by 87%, is virtually the same and there is a substantial increase in total output and output per worker in both counterfactuals. We discuss the wage-inelastic equilibrium results, noting any differences from the Lewis equilibrium.

The most notable change from expanding the size of farms to the technology-conditional optimum of 24 acres, from the average size of 3.1 acres, is that there is an expansion of total output by 42% and, due to the substantial reduction in the size of the agricultural labor force by 16%, an increase in per-worker output of over 68%, a change that is only 4% lower than the increase in the Lewis equilibrium in which wages stay constant. There are thus both too many farms and too many persons engaged in agricultural sector in terms of the return on land (profitability per acre) and in terms of total output, with the Indian rural work force as a whole substantially hurt by the small size of Indian farms as measured by output per worker. However, if there is not a Lewis equilibrium, landless laborers, who make up a substantial fraction of the agricultural labor force, are made slightly worse off due to the wage decline in the absence of any redistributions of the output gains. Absorption of the rural labor force into the industrial/service sector is thus key to achieving the gains in agriculture that does not penalize a substantial fraction of the population.

What are the sources of the gains? Farm efficiency, measured as profits per acre, of course increases at the optimal farm size, by 13%. The key reason for the substantial rise in output is that the larger farms permit the use of higher capacity machines. This lowers the marginal cost of providing plant nutrients to the farm and thus leads to a substantial increase in output. All the

optimal-sized farms use machinery, compared to only 21% of farms currently, with mode machine capacity growing by 72%. Because machinery and workers are substitutable, this leads to a reduction in the workforce.

Perhaps what is surprising is that the workforce declines, while substantial, are not larger than they are given the elasticity of substitution of 13.9 between worker and machine use. The key reason that the workforce does not decline substantially due to higher machine use - machine hours per acre per day expands from 0.71 at 7 acres (below this level the model predicts that machines are not used) to 3.8 at 24 acres - is that labor is significantly underutilized in the current regime. This is because of the existence of fixed labor costs. As noted, because of these costs, a substantial fraction of the acreage in Indian agriculture is under autarchy. And on these farms farm labor is limited to the family such that the marginal product of labor exceeds the marginal cost of labor, inclusive of fixed hiring costs, on other farms by 42% on average. Our calibrated model indicates that labor hours per acre per day on a 24 acre farm is 4.75 and that on a one-acre farm, in which family labor works off farm, is 9.7. But on 9-acre autarchic farms, the threshold acreage above which hired labor is employed, predicted labor hours per acre per day is only 2.7. We noted above that 34% of farms in India are in autarchy, but since autarchic farms are larger than the mean farm, 52% of the acreage is currently farmed in autarchy. Thus, current labor utilization in Indian agriculture is substantially less than it would be in the absence of the fixed costs of hiring due to the substantial fraction of acreage that is cultivated under autarchy, and no farms are in autarchy in the regime with 24-acre farms.

5. Conclusion

Much of the recent literature on economic development in agriculture has focused on the adoption of new technology. The thrust of the argument is that productivity differences across countries are importantly driven by the fact that farmers in low income countries do not have access

to the same types of seeds and inputs that are available in other countries. One of the salient differences in agricultural sectors between countries of very different productivities, however, the scale of operations, has been relatively ignored.⁴⁶ The evidence on persistent differences in productivity within countries associated with scale cannot be easily understood to be a consequence of differential lack of access to technologies given evidence on the transmission of technologies between farmers over time. If farmers within an area have equal access and markets for inputs work reasonably well, marginal products should be equated across farmers and thus it is unclear why there should be any differences in productivity across farms. Moreover, despite the global differences in farm productivity indicating higher productivity among larger farms, almost all of the evidence from within low-income countries suggests that agricultural productivity and scale are inversely related over the full distribution of farm sizes observed in those countries. Understanding what lies at the source of these differentials by scale may thus be indicative of important barriers to development in rural areas of low-income countries.

In this paper we examined unique data from India that allows us to look at agricultural operations among a wider distribution of farm scales than is typically observed in low-income countries because of the over-sampling of larger farms. We found a distinctive U-shaped pattern in which both small and large farmers are more productive, in terms of both yield and profitability, than intermediate sized farmers. This pattern replicates within one rural setting what is observed across countries, with productivity decreasing in scale for smaller farms and increasing in scale for larger farms. We showed that these productivity patterns by scale are not attributable to differences in measured aspects of land quality. We also established that area measurement error, that might lead to high profitability per acre on small farms, is small and does not importantly explain the observed

⁴⁶Seed technologies across low-income and developed-country agriculture are in fact not that different. Most of the cotton grown by the ICRISAT farmers is modern BT cotton, with cotton being the second most cultivated crop by area planted. The Indian farmers are also using 'Roundup Ready' soybeans, the same as used in the United States and Brazil.

patters. In addition we found that the U-shape pattern is observed across plots for the same farmer, thus ruling out credit access as the main explanation of higher profitability among large farmers.

We proposed two complementary mechanisms that drive productivity differences by scale and can account together for the U-shape pattern by scale across countries of the world and in India. We first considered the role of fixed costs in the hiring of labor and other inputs. We provided evidence that in fact many workers work for less than a full day and that, consistent with the existence of fixed transaction costs to hiring labor, the hourly price of a worker is higher when workers are used for part of day, a large proportion of farms and agricultural operations are in autarchy - with neither family workers working off the farm nor any laborers hired - and that intermediate sized farmers are most likely to employ workers part time. The implication of this pattern is that small farmers will be relatively efficient because the shadow price of family worker time is set by the outside market, as most smallholders work part-time off farm. As farm size grows however, the farm moves to autarchy and is reluctant to take on hired workers for just a few hours per operation. The underutilization of labor leads to lower yields and profitability per acre. Eventually, this strategy proves costly and there is discrete jump upward in total work per acre due to the hiring of non-family workers.

Transaction costs in hired labor can explain the declining relationship between size and productivity and the existence of autarchic agricultural production characterizing most low-income countries, but cannot explain the observed continuous rise in profitability by scale beyond a threshold of farm size. The second mechanism we focus on to explain this component of the U-shape is the adoption of machine technology that is differentially adapted to farm size. The idea is that higher capacity machines do more work per hour and that the cost of these higher machines does not increase proportionately with capacity. However, large machines cannot be used at full capacity on small farms or plots. As a result large farmers use more productive machines and small

farmers, if they use machines at all, use less productive machines that are also less cost-effective. This leads to an increase in yield and profitability as farm size increases.

We focused in particular on power sprayers because we are able to measure capacity based on time use and the amount that is sprayed. We showed that sprayers fit well with our model - optimal sprayer capacity increases with farm size and the shadow price of sprayer capacity increases less than proportionally as capacity increases. The mechanisms of hired-labor transaction costs and economies of scale in machine capacity together allowed us to replicate the patterns of agricultural profitability and input costs per acre by farm size that are observed in our data and across countries of the world. Based on our parameter estimates, we found in particular that in our sample, economies of scale peter out at the very top of the land distribution (24.1 acres), suggesting that average farm size in India is, along this one dimension of technology that we can precisely estimate, too small - there are too many farms.⁴⁷

We calibrated our model, in which neither TFP nor the production technology exhibits any scale economies, to show that labor market transaction costs and machine capacity scale economies are able by themselves to capture both the levels and non-linear patterns of aggregate measures of farm outputs and inputs by acreage in the VLS study areas of rural India. Calibration of the model also yields estimates of the transaction costs of hiring labor that are consistent with the wage and hours schedules in the data and reveal that for the 34% of farm operations (accounting for over 50% of total land in India) in autarchy, the marginal opportunity cost of labor (marginal labor product) is on average over 40% higher than it is on the smallest and largest farms. Thus, the existence of high

⁴⁷The optimality of 24 acres is conditional, as noted, on the machine technology available in India. There is machinery that exists for much larger farms in the world - sprayer drones, for example. Such technology is not widely available in India because there are no farms of sufficient scale that could profitably exploit it. Machine technology available in India is endogenous to farm size. Another reason that 24 acres is not the true limit in India for maximizing surplus is that we have assumed that even on 24-acre farms the episodic nature of farm operations, and the heterogeneity of labor demand across operations for a given acreage, makes hiring labor on long-term contracts to minimize labor market transaction costs infeasible. But with larger farms, relational contracting may result in additional cost savings.

labor market transaction costs leads to a substantial mis-allocation of labor across farms as well as the underutilization of rural labor on a large fraction of farms

To assess the cost of the existing structure of land and labor allocations in India in terms of total agricultural output and output per agricultural laborer we used our calibrated model to carry out a counterfactual in which we eliminated all scale differences across farms and fixed farm size at the level such that, by exploiting the scale economies of available machines in India, maximizes profitability per acre. We embedded the model in an equilibrium framework in which workers can migrate to urban areas, where labor demand is inelastic. The new calibrated equilibrium, in which the optimal farm size at the equilibrium wage rate and conditional on existing machine technology, is 24 acres, compared with a mean of just over three acres in India today, results in a reduction in the number of farms from over 95 million to 12.4 million.

This reduction in the number of farms, which eliminates the mis-allocation of labor across farms and permits exploitation of the scale economies arising from machinery that lowers the marginal cost of agricultural inputs, results in a 50% increase in agricultural output. It also results in a reduction in the rural workforce. There is indeed surplus labor, as suggested in the writings of Arthur Lewis and others, as well as a surfeit of farms. However, despite the high degree of evident substitutability between workers and machines, the large reduction in the number of farms leads to only a 16% decline in the labor force and a 9% decline in the rural wage, due to the current underutilization of labor in the existing rural economy associated with autarchic production. The resulting decline in the rural labor force and the rise in output per-acre from the simulation reducing the number and heterogeneity of farms in India leads to substantial increase in total output per worker of 68%.

The counter-factual calibration exercise of increasing the size of farms in India is meant to provide estimates of the costs of the current distribution and small scale of agriculture in India given

existing machine technology in India. Carrying out a policy that reduced the number of farms in India would have a number of major challenges. While expanding the urban labor force by 22% (which is what would be required to accommodate these workers), involves a considerable amount of disruption for a great number of people, it does not seem prohibitive given that there was an expansion in main workers in urban areas of 44% over the period from 2001 to 2011. But perhaps the more difficult issue would be the implied transition of a large number of current small-scale cultivators to, albeit less poor if compensated, landless hired workers as the number of farms declines by 87%.

Our findings also suggest that an endogenous evolution of the distribution and number of farms India to resemble those attributes of farms in high-income countries, even barring cultural, economic, or legal barriers to the buying or selling of land, is unlikely given the U-shaped profitability curve. Any marginal increase in acreage for most of the farms in India, who are located on the declining portion of the output-per-acre curve, would reduce profitability per acre. As a consequence, the selling price of the marginal acre faced by a buyer who owns a small farm would exceed the discounted lifetime flow of profits from that additional acreage. Moving from the mean farm size of three acres to a more profitable farm on the rising segment of the productivity curve would thus require multiple transactions, given the generally small size of farms. The large number of transactions would be complicated even in a relatively competitive market, but the need for contiguous plots raises important issues such as hold-up problems that will fully extract the rents that would otherwise accrue to a farmer who puts together a large farm.⁴⁸

In sum, despite the potential gains to farm productivity and output per worker of transitioning to an equilibrium of fewer but larger farms, small farms are likely to be the dominant

⁴⁸The same issues would appear applicable to farmers cooperating among themselves to exploit scale economies by operating their farms jointly - many farmers would have to coordinate production, not just one or two, and all would have to be from contiguous farms.

force of production in low income countries for the foreseeable future without external intervention.⁴⁹ Our analysis suggests, however, as partly envisioned by Arthur Lewis, an expansion of the industrial sector that draws workers out of the agricultural sector would not only not reduce output in the agricultural sector but via the reduction in the number of farms and the concomitant increase in farm scale would increase total agricultural output and rural per-capita income.

⁴⁹In recent years Brazil has experienced a substantial transition to large-scale agricultural production in many areas. This reflects the availability of large-scale machines, the conversion of land that had previously been used for low intensity activities such as grazing, and the development of an effective series of highly profitable export markets.

References

- Ali, D. A, and K. Deininger. 2014. "Is there a Farm-Size Productivity Relationship in African Agriculture? Evidence from Rwanda." World Bank Policy Research Paper 6770, World Bank, Washington, DC.
- Allen, Robert C. 1988. "The Growth of Labor Productivity in Early Modern English Agriculture." *Explorations in Economic History* 25, 117-146.
- Assunção, Juliano J., and Luis H. B. Braido. 2007. "Testing Household-Specific Explanations for the Inverse Productivity Relationship." *American Journal of Agricultural Economics* 89 (4): 980-90.
- Barrett, C. B., M. F. Bellemare, and J. Y. Hou. 2010. "Reconsidering Conventional Explanations of the Inverse Productivity-Size Relationship." *World Development* 38 (1): 88-97.
- Benjamin, D. 1995. "Can Unobserved Land Quality Explain the Inverse Productivity Relationship?" *Journal of Development Economics* 46 (1): 51-84.
- Bhalla, S. S., and P. Roy. 1988. "Misspecification in Farm Productivity Analysis: The Role of Land Quality." *Oxford Economic Papers* 40 55-73.
- Binswanger-Mkhize, H.P, Camille Bourguignon, and Rogier van den Brink. 2009. *Agricultural Land Redistribution: Toward Greater Consensus*. Washington D.C.: World Bank.
- Binswanger, H.P., K. Deininger, and G. Feder. 1995. "Power Distortions Revolt and Reform in Agricultural Land Relations." In J. Behrman, and T.N. Srinivasan, ed. *Handbook of Development Economics*, Volume III. Amsterdam, the Netherlands: Elsevier Science B.V.
- Bivar, Venus. 2010. *The Ground Beneath Their Feet: Agricultural Industrialization and the Remapping of Rural France, 1945-1976*. Department of History, University of Chicago Dissertation.
- Cai, Zongwu, Mitali Das, Huaiyu Xiong, and Xizhi Wu. 2006. "Functional Coefficient Instrumental Variables Models." *Journal of Econometrics* 133 (1): 207-241.
- Carletto, C., S. Savastano, and A. Zezza. 2013. "Fact or Artifact: The Impact of Measurement Errors on the Farm Size Productivity Relationship." *Journal of Development Studies* 103: 254-261.
- Carter, Michael. R. 1984. "Identification of the Inverse Relationship between Farm Size and Productivity: an empirical Analysis of Peasant Agricultural Production." *Oxford Economic Papers* 36: 131-45.
- Carter, M. R., and K. Wiebe. 1990. Access to Capital and its Impact on Agrarian Structure and Productivity in Kenya. *American Journal of Agricultural Economics*, 72: 1146-1150. 41
- Collier, P., and S. Dercon. 2014. "African Agriculture in 50 Years: Smallholders in a Rapidly Changing World?" *World Development* 63: 92-101.
- Davis, C. Austin. 2016. "Why Did Sugarcane Growers Suddenly Adopt Existing Technology?" University of Michigan.
- Foster, A., and M. R. Rosenzweig. 1996. "Comparative Advantage, Information and the Allocation of Workers to Tasks: Evidence from an Agricultural Labour Market." *The Review of Economic Studies* 63(3): 347-374.
- Foster, A., and M. R. Rosenzweig. 2011. *Are Indian Farms too Small? Mechanization, Agency Cost, and Farm Efficiency*. New Haven CT: Economic Growth Center, Yale University.
- Goldar, Bishwanath (2009). "Trade Liberalisation and Labour Demand Elasticity in Indian Manufacturing." *Economic and Political Weekly* 44 (34):51-57.
- Hayami, Yujiro, and Keijuro Otsuka. 1993. *The Economics of Contract Choice: An Agrarian Perspective*. Oxford: Oxford University Press.

- Hazell, Peter. 2011. "Five Big Questions about Five Hundred Million Small Farms." Paper presented at the IFAD Conference on New Directions for Smallholder Agriculture, IFAD Headquarters, Rome, Italy, 24-25 January, 2011.
- Hazell, P., C. Poulton, S. Wiggins, and A. Dorward. 2010. "The Future of Small Farms: Trajectories and Policy Priorities." *World Development* 38(10): 1349-1361.
- Hornbeck, Richard and Suresh Naidu. 2014. "When the Levee Breaks: Black Migration and Economic Development in the American South." *American Economic Review* 104(3): 963–990.
- Huang, Jikan; Ruifa Hu; Jianmin Cao; and Scott Rozelle. 2008. "Training Programs and In-the-field Guidance to Reduce China's Overuse of Fertilizer Without Hurting Profitability." *Journal of Soil and Water Conservation* 63 (5): 165A-175A.
- Jayne, T.S., J. Chamberlin, and Derek D. Headey. 2014. "Land pressures, the evolution of farming systems, and development strategies in Africa: A synthesis." *Food Policy*, 48: 1-17.
- Kagin, J., E. Taylor, and A. Yúnez-Naude. 2015. Inverse Productivity or Inverse Efficiency? Evidence from Mexico, *Journal of Development Studies* DOI:10.1080/00220388.2015.1041515.
- Kimhi, A. 2006. "Plot Size and Maize Productivity in Zambia: Is there an Inverse Relationship?" *Agricultural Economics* 35(1): 1-9.
- Larson, D., K. Otsuka, T. Matsumoto, and T. Kilic. 2013. "Should African Rural Development Strategies Depend on Smallholder Farms? "An Exploration of the Inverse Productivity Hypothesis." *Agricultural Economics* 45: 1-13.
- Lewis, W. Arthur (1954). "Economic Development with Unlimited Supplies of Labor". *The Manchester School* 22 (2): 139–91.
- Lichter, Andreas, Andreas Peichl and Sebastian Sieglöch (2015). "The own-wage elasticity of labor demand: A meta-regression analysis." *European Economic Review* 80: 94-119.
- Mazumdar, D. 1965. "Size of Farm and Productivity: A Problem of Indian Peasant Agriculture." *Economica* 32(126): 161-173.
- Muyanga, Milu and T.S. Jayne. 2019. "Revisiting the Farm Size-productivity Relationship Based on a Relatively Wide Range of Farm Sizes: Evidence from Kenya." *Amer. J. Agr. Econ.* 101(4): 1140–1163.
- Catherine Morrison Paul, Catherine Morrison; Catherine Paul; Richard Nehring; David Banker, and Agapi Somwaru. 2004. "Scale Economies and Efficiency in U.S. Agriculture: Are Traditional Farms History?" *Journal of Productivity Analysis* 22(3): 185-205 .
- Schultz, T.W. 1964. *Transforming Traditional Agriculture*. New Haven, CT: Yale University Press.
- Stiles, Scott and C. Robert Stark, jr. 2016. "Estimating Farm Machinery Costs." *FSA21*, Division of Agriculture, Research and Extension, University of Arkansas.
- United Provinces Zamindari Abolition Committee. 1948. *Report Volume 1*. Allahabad, U.P., India.
- Vollrath, D. 2007. "Land Distribution and International Agricultural Productivity." *American Journal of Agricultural Economics* 89(1): 202-216.
- Yotopoulos, P.A., and L.J. Lau. 1973. "A Test for Relative Economic Efficiency: Some Further Results." *Am. Econ. Rev.* 63(1): 214–223.
- Zaibet, L.T., and E.G. Dunn. 1998. "Land Tenure, Farm Size, and Rural Market Participation in Developing Countries: The Case of the Tunisian Olive Sector." *Economic Development and Cultural Change* 46(4), 831-848.
- Zhang, B., and A.C. Carter. 1997. "Reforms, the Weather, and Productivity Growth in China's Grain Sector." *American Journal of Agricultural Economics*, 79(4): 1266-1277.

Appendix A1: Measurement Error, Plot Quality and Farmer Ability

To assess the role of measurement error in determining the U-shaped relationship between farm productivity and scale, we exploit the fact that we have two independent, and different, elicitation of farm size in 2009. Each household was visited and administered a Census questionnaire asking for the total number of acres of owned unirrigated and owned irrigated land. At a separate time in the same crop year, the panel households were asked to provide the acreage of each individual plot they owned. Thus, there are two sets of information on acreage for each panel household collected at different times. We summed the individual plot acreages to obtain total farm size from the survey and summed the reported total acreages of irrigated and unirrigated land to obtain total farm size from the Census elicitation. We then used the latter as an instrument for the former to estimate the effect of farm size on profits. Appendix Table A1 reports the OLS and IV profit estimates. The specifications also include dummy variables for each village, capturing the influences of weather and prices at the village level. As can be seen, the IV and OLS estimates are economically identical. However, the fundamental issue is not how much measurement error there is in acreage reports, but whether measurement error leads to bias that is correlated with farm size. For example, even if measurement error is on average small, it may be proportionately large for small plot sizes, which would create a larger negative bias in the relationship between size and per-acre profitability at small compared with big farms. To assess this possibility, we obtained the coefficient of land size by land size, using the locally-weighted functional coefficient model (LWFCM),⁵⁰ obtained again using OLS and IV. Figure A2 displays the two land coefficients by land size, indicating that measurement error is small at all acreages. Measurement error does appear to increase with farm size above ten acres, suggesting that the upward component of the U-shape for profits per acre displayed in Figure A2 may actually be slightly understated due to measurement error. Measurement error, however, is clearly not the major cause of the downward slope in profitability below 10 acres.

The U-shape also survives control for land quality. The survey provides, at the plot level, multiple measures of plot characteristics. These include soil depth, distance of the plot from the household residence, four categories of soil fertility, six levels of soil degradation, and 11 soil types. Appendix Table A2 reports the relationship between real output and real *Kharif* profits and farm size for the full sample. In the first column for each dependent variable, the specification only includes village/year fixed effects. The second column estimate for each is from a specification additionally including all the plot characteristics. The difference in estimates is negligible, despite the fact that the set of 24 variables associated with the plot characteristics are highly statistically significant for both profits and gross output value.⁵¹

We can also rule out credit constraints and farmer ability as the sole determinants of the U-shape relationship. To assess whether credit constraints associated with farmer characteristics, such as total wealth or ability, are responsible for the U-shape in the third and fourth columns of Table A2 for each dependent variable we use information on output and profits at the plot level. The third column for each reports the within-farmer-year estimate of plot size; the fourth column reports the within-farmer-year estimate of the plot size effect including the soil characteristics of the plots. Again, the estimates are virtually identical. The within-farmer-year plot size estimates are lower than the cross-farmer farm size estimates, suggesting that farmer ability or wealth may play some role in

⁵⁰See Cai *et al.*, 2006. The specification we use is locally linear in profits and farm size.

⁵¹Below, we will provide plot fixed-effects estimates of operational scale effects by plot size that control completely for land quality.

the relationship between farm size and productivity, but this form of heterogeneity is not responsible for the observed scale effects.

Appendix A2

Assume the smoothed fixed cost functions and evaluate profits per unit area at the acreage where family workers are working off-farm (point 1) and just at the cusp between off-farm work and autarchy (point 2). Then if π_1 and o_1 denote profit per unit and area and output per worker, respectively, at point 1 and so forth then we may define:

$$\pi_1 = \psi a_1^\alpha l_{f1}^{1-\alpha} - \frac{w_0^1 l_{f1}}{8} - w_1 l_{f1}$$

$$\pi_2 = \psi a_2^\alpha l_{f2}^{1-\alpha} - \frac{w_0^2 l_{f2}}{8} - w_1 l_{f2}$$

$$o_1 = \psi a_1^\alpha l_{f1}^{1-\alpha} / l_{f1}$$

$$o_2 = \psi a_2^\alpha l_{f2}^{1-\alpha} / l_{f2}$$

Given l , a_1 , a_2 and the profit and output measures at these points

$$\psi = \frac{l^{-1+\alpha} a_2^{-\alpha} \pi_2}{1 + l^\alpha (-1 + \alpha) o_2^{-\alpha}}$$

$$w_0^2 = \frac{-8o_2^\alpha w_1 l - 8(-1 + \alpha)(l^{\alpha+1} w_1 + \pi_2 l^\alpha)}{((-1 + \alpha)l^\alpha + o_2^\alpha)l}$$

$$\alpha = -\ln\left(\frac{\psi}{o_1}\right)\left(\mathbb{W}\left(-\frac{o_1 a_1}{\pi_1} \ln\left(\frac{\psi}{o_1}\right)\right)\right)^{-1}$$

$$w_0^1 = \left((-8w_1 + 8o_1) \mathbb{W}\left(-\frac{o_1 a_1}{\pi_1} \ln\left(\frac{\psi}{o_1}\right)\right) + 8o_1 \ln\left(\frac{\psi}{o_1}\right) \right) \left(\mathbb{W}\left(-\frac{o_1 a_1}{\pi_1} \ln\left(\frac{\psi}{o_1}\right)\right) \right)^{-1}$$

where $\mathbb{W}()$ denotes Lambert's W function. Then

$$n_0 = \frac{w_0^1 - w_0^2}{a_1 - a_2}$$

and

$$w_0 = w_0^1 - n_0 a_1.$$

Figure 1. Percent of Households with Operational Landholdings Below 10 Acres, by Country

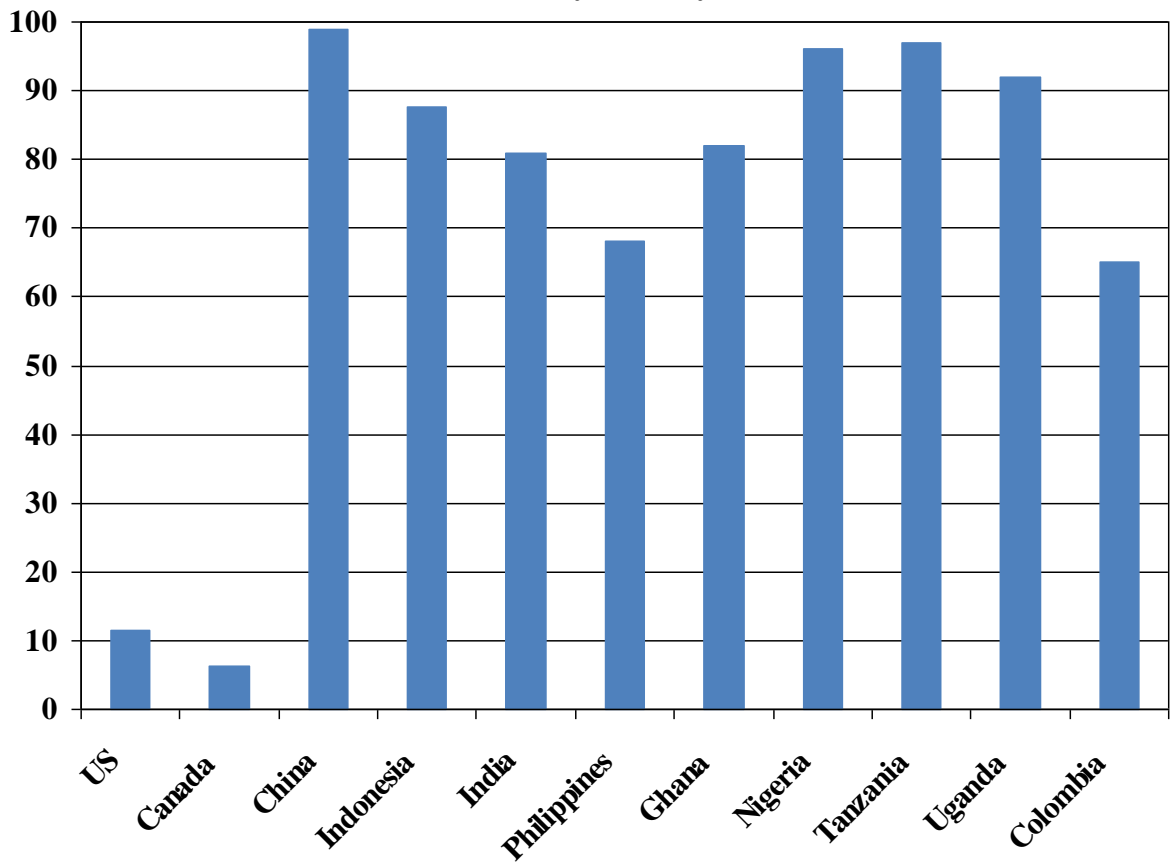


Figure 2. Soybean Yields (Metric Tons per Hectare) in 2016, by Country
(Source: USDA, 2016)

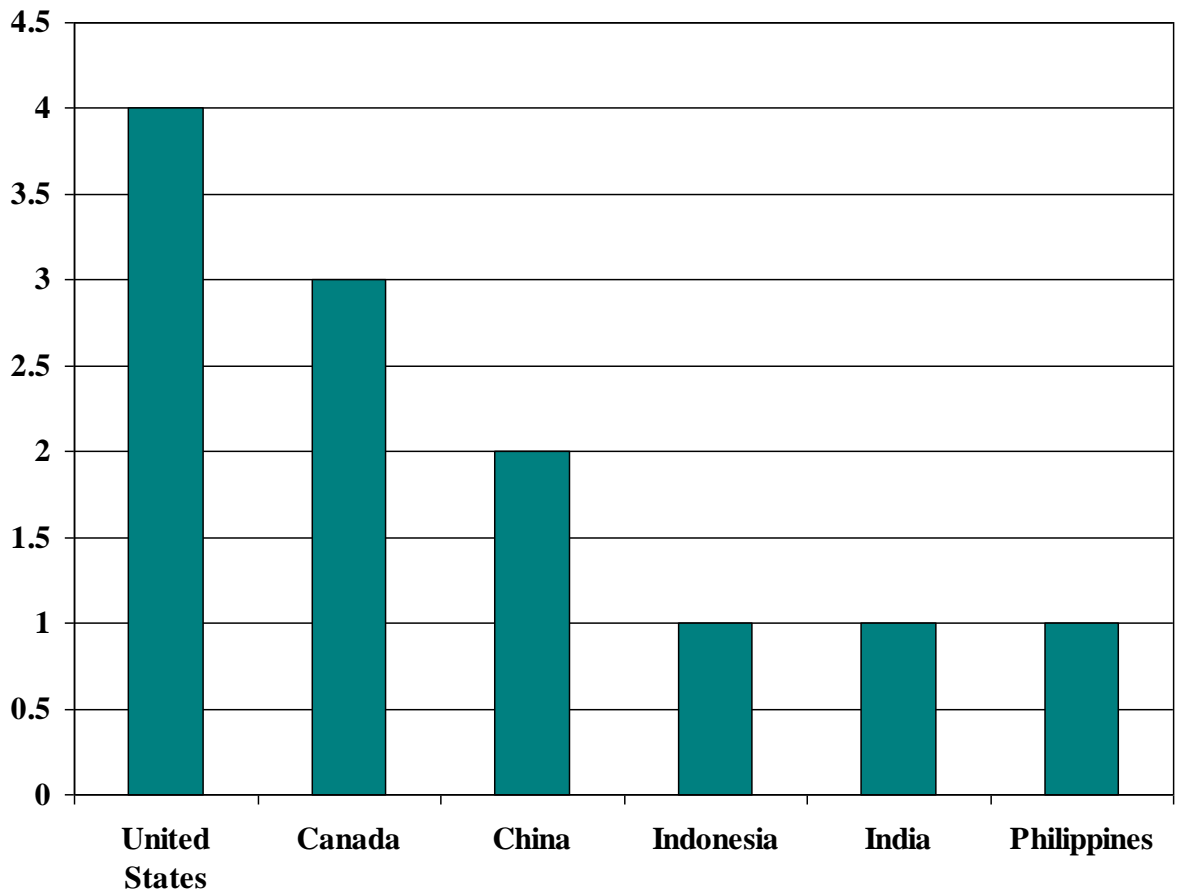
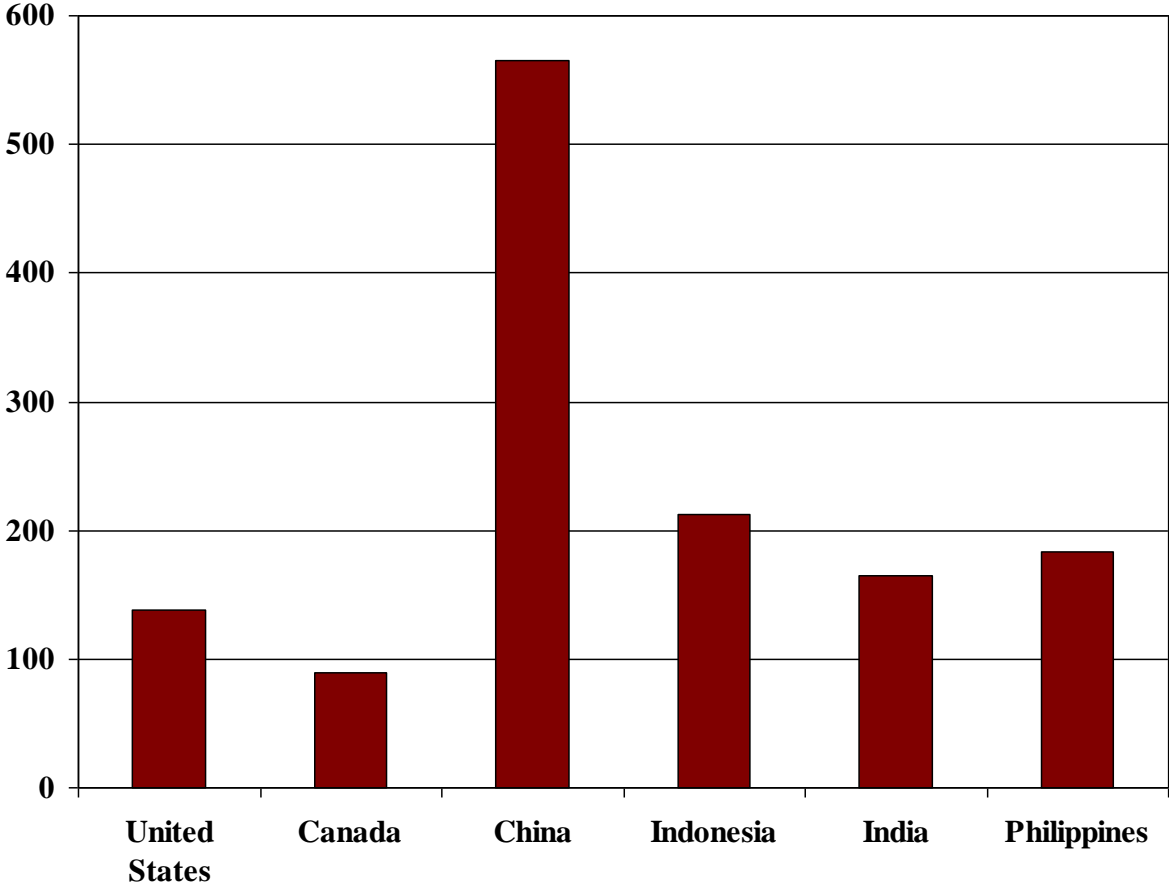
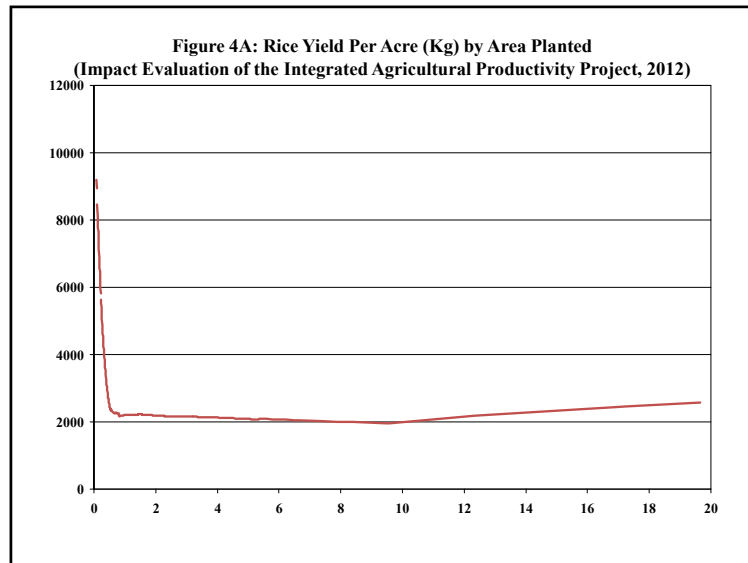
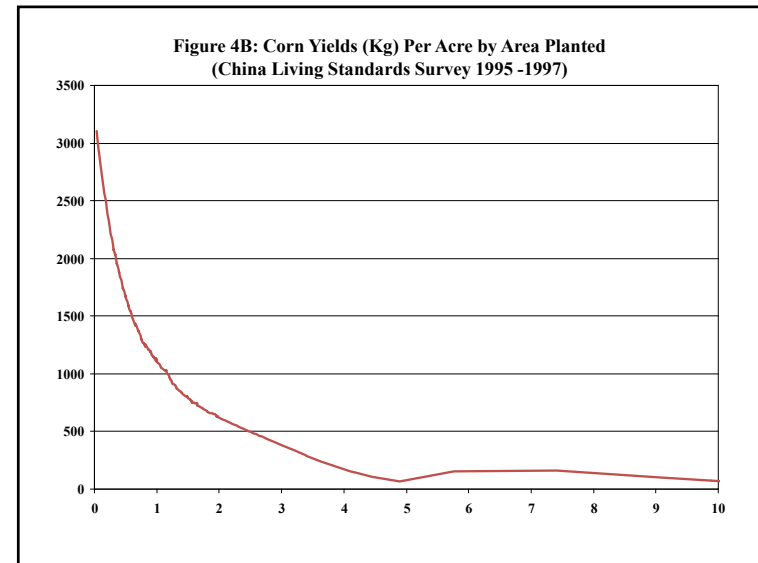


Figure 3. Fertilizer Intensity (Kilograms per Hectare) in 2014, by Country
(Source: World Bank, 2016)

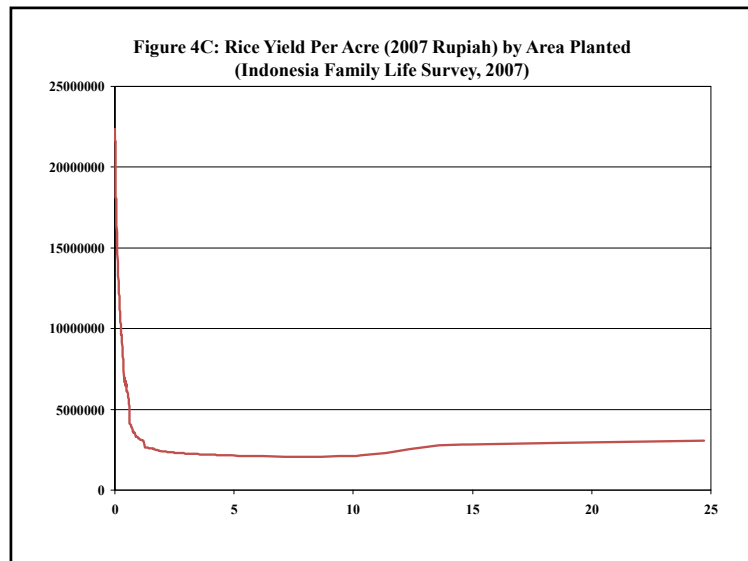




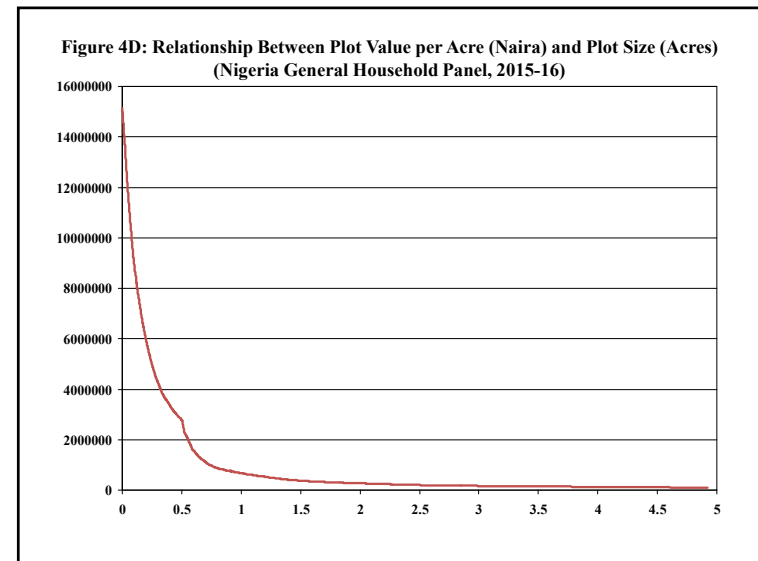
1



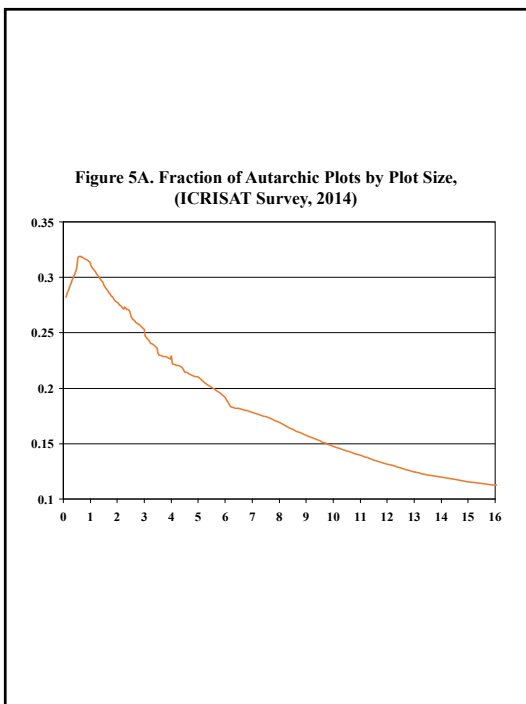
2



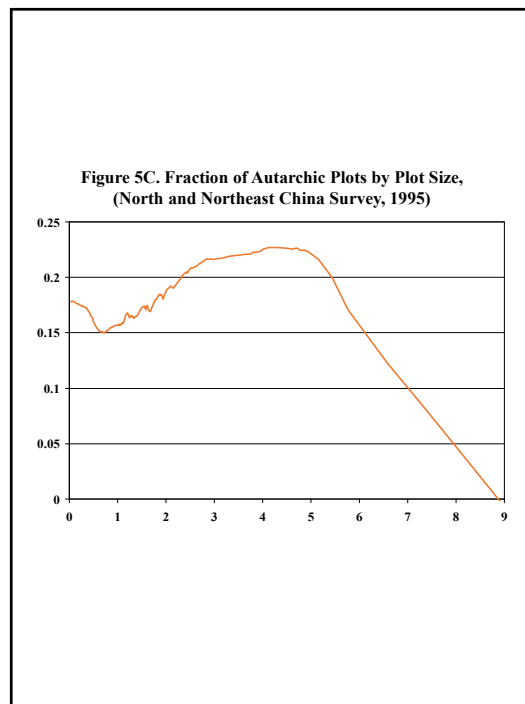
3



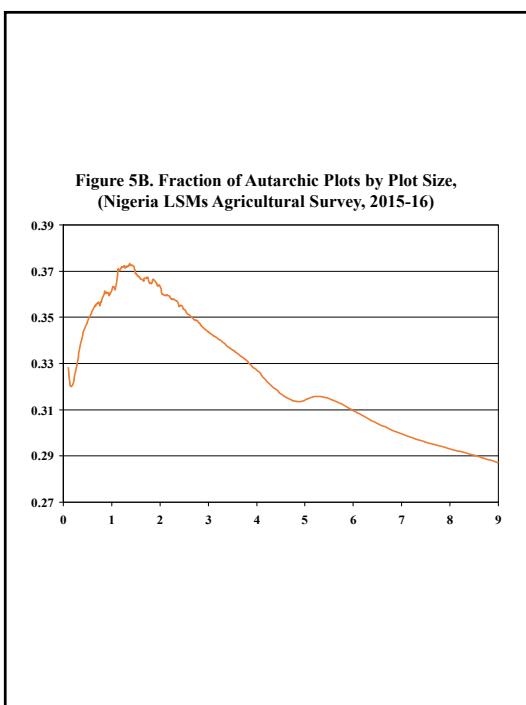
4



1



3



2

**Figure 6. Cumulative Distributions of Owned Total Land and Land Plots (Acres),
By Sample and Census: ICRISAT VLS 2014**

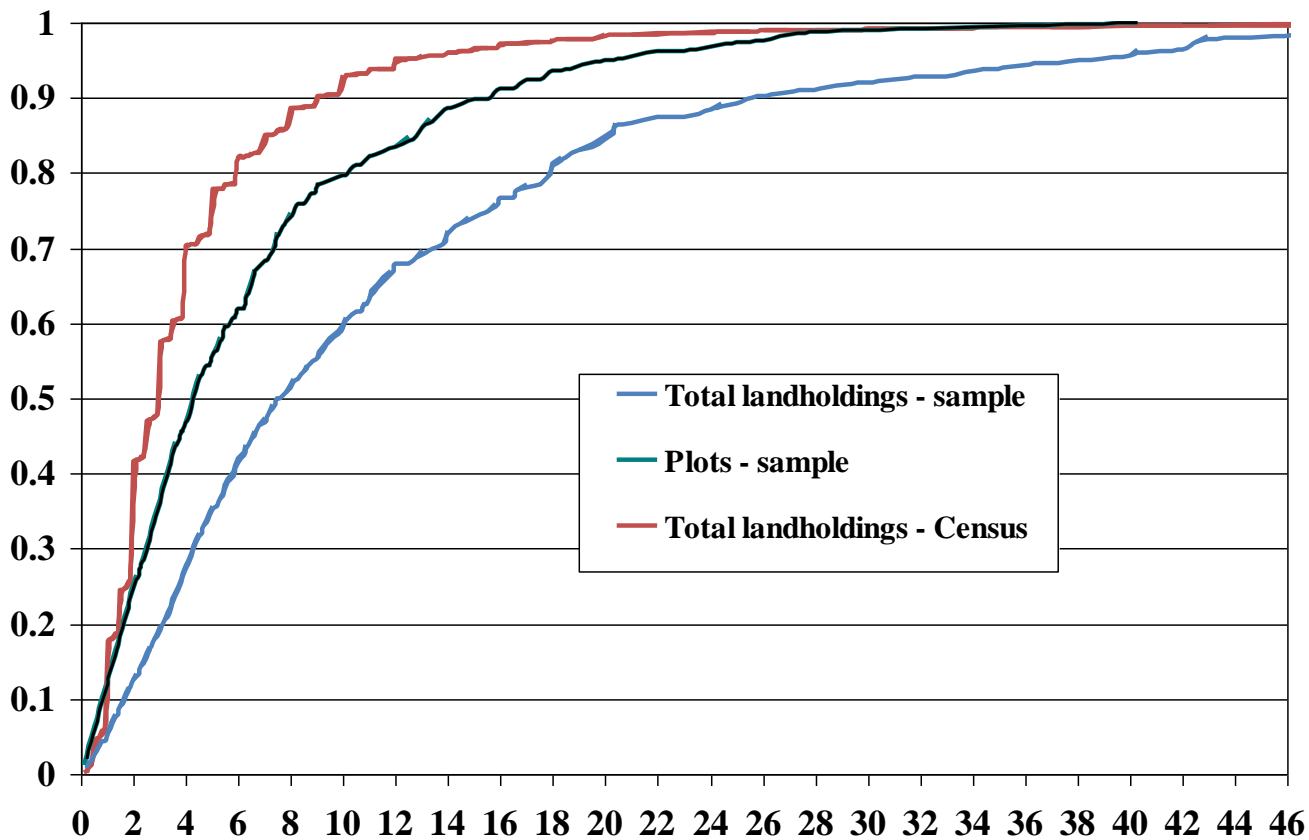


Figure 7. Relationship Between Real Average Profits per Acre and Farm Size (Acres)
(ICRISAT VLS 2009-14)

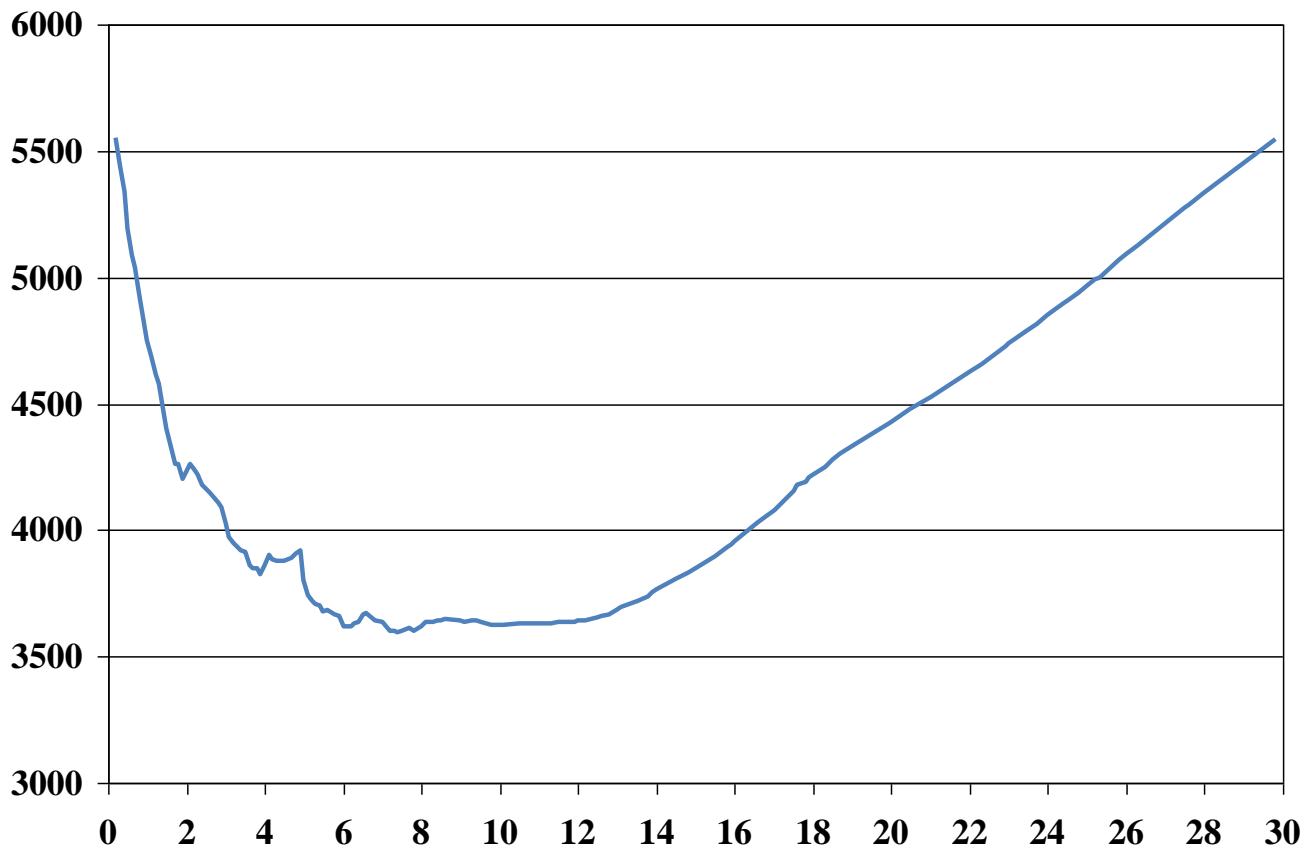


Figure 8. Real Profits per Acre by Owned Area: Roles of Plot Quality and Farmer Characteristics (ICRISAT VLS 2009-14)

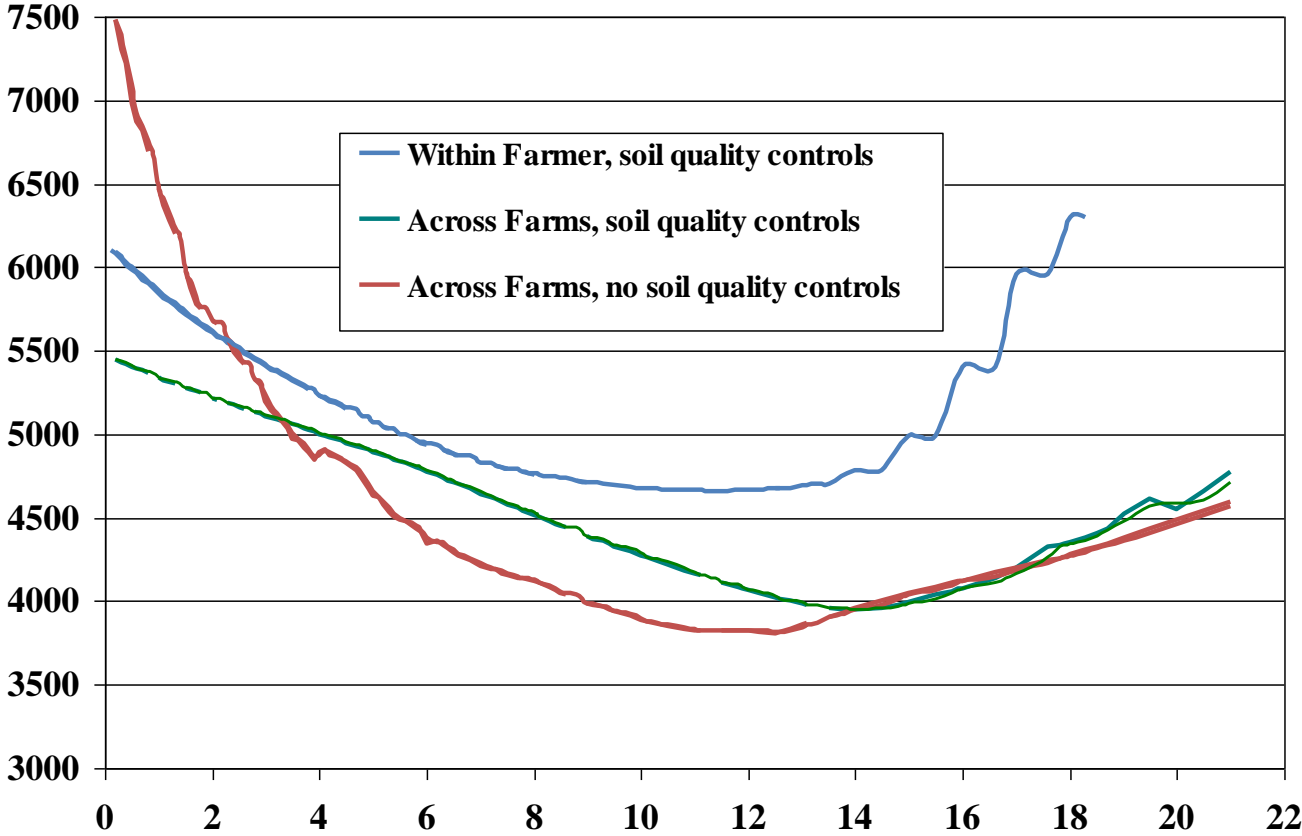


Figure 9: Input Supply: Distribution of Average Hours Worked per Day for Wages, *Kharif* Season, by Hired Input

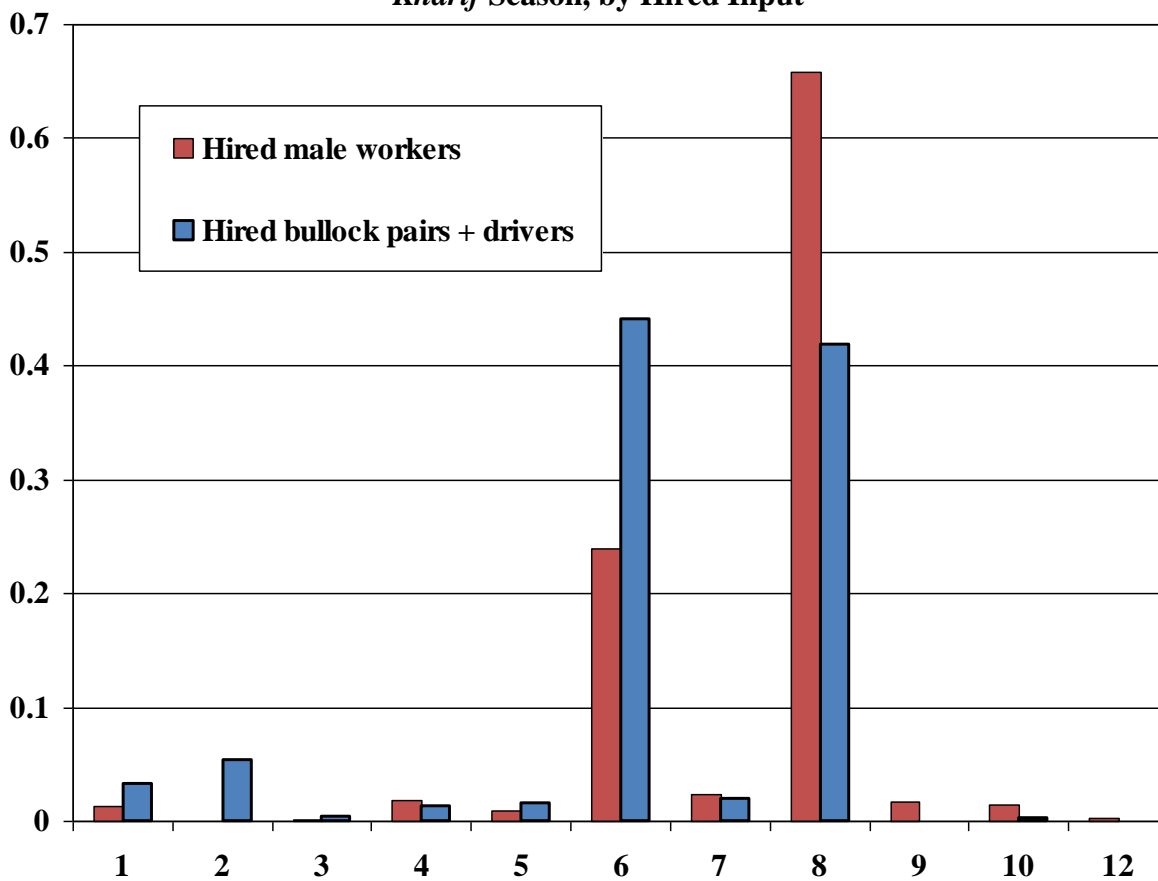
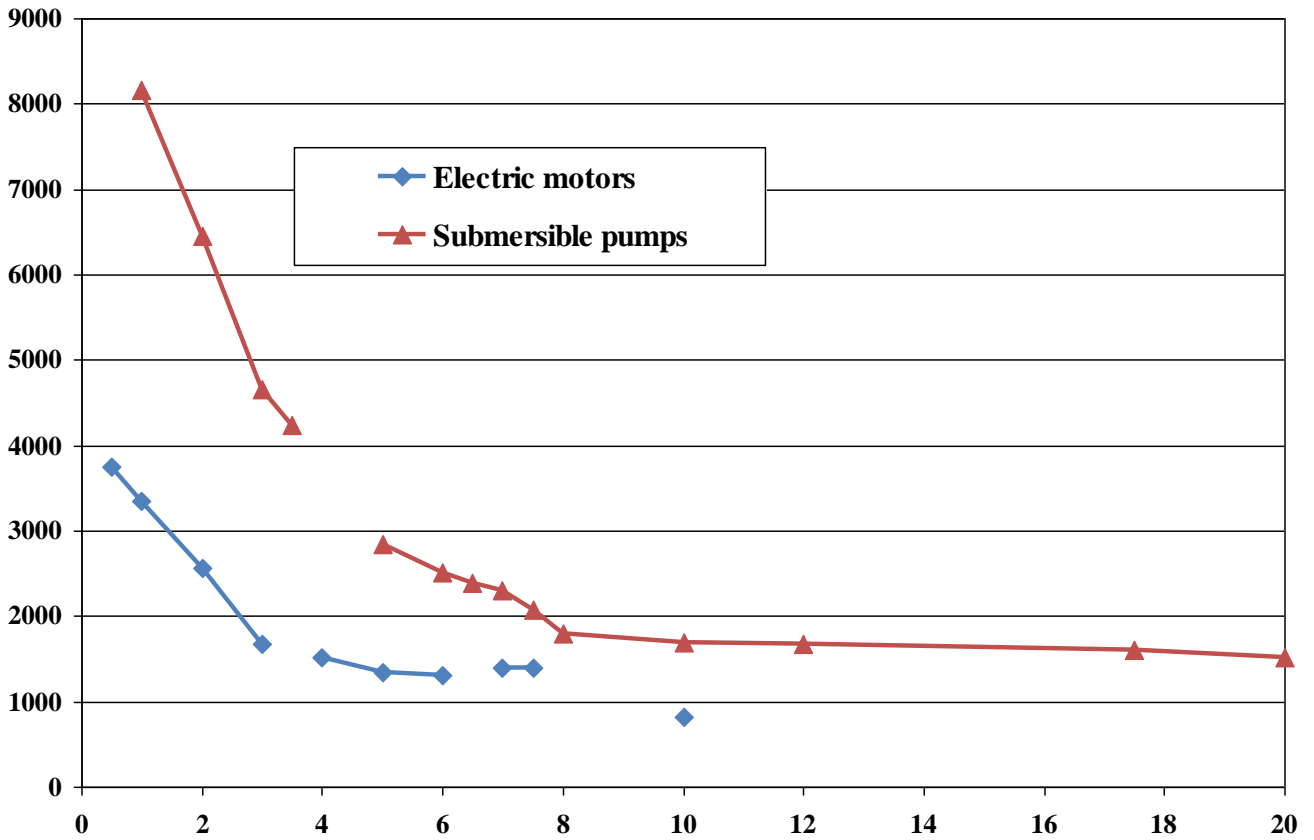
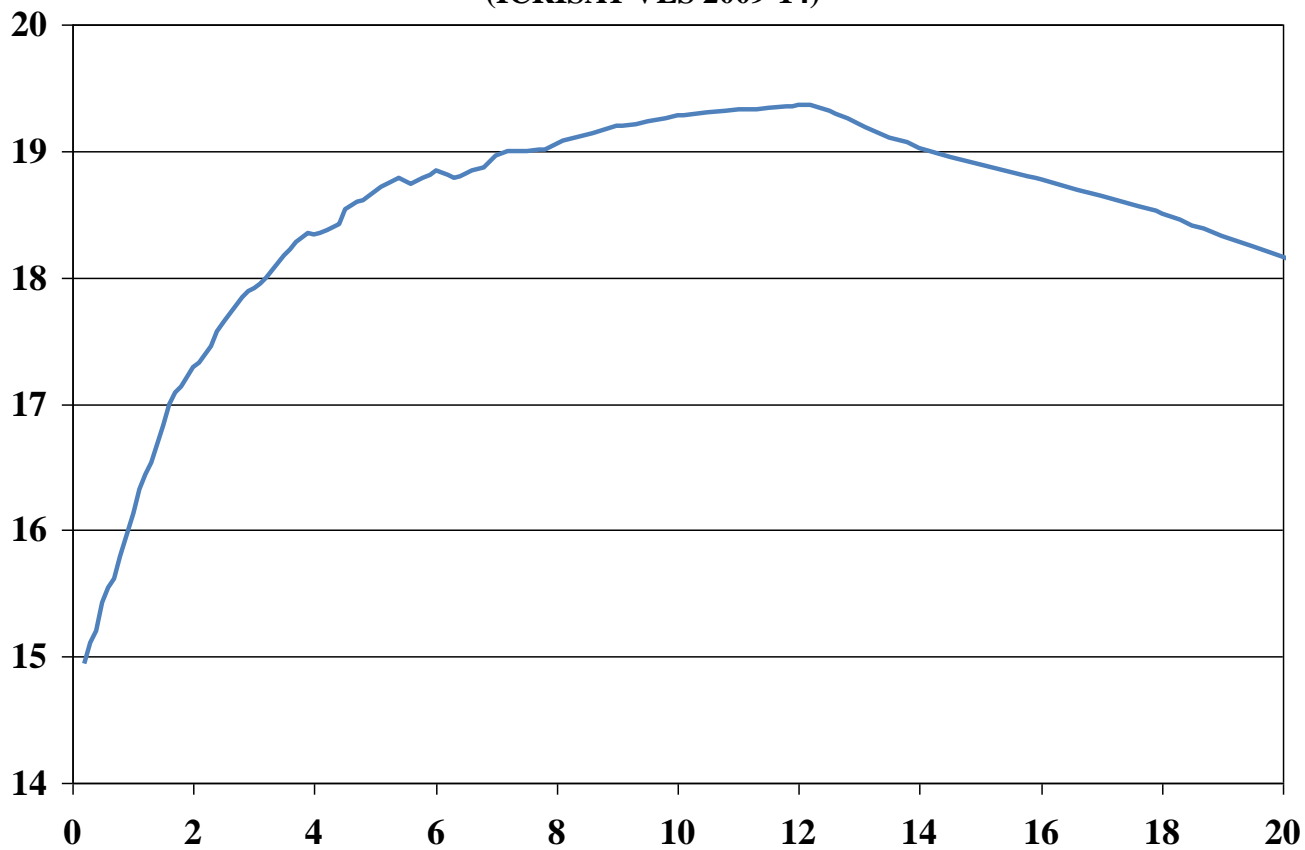


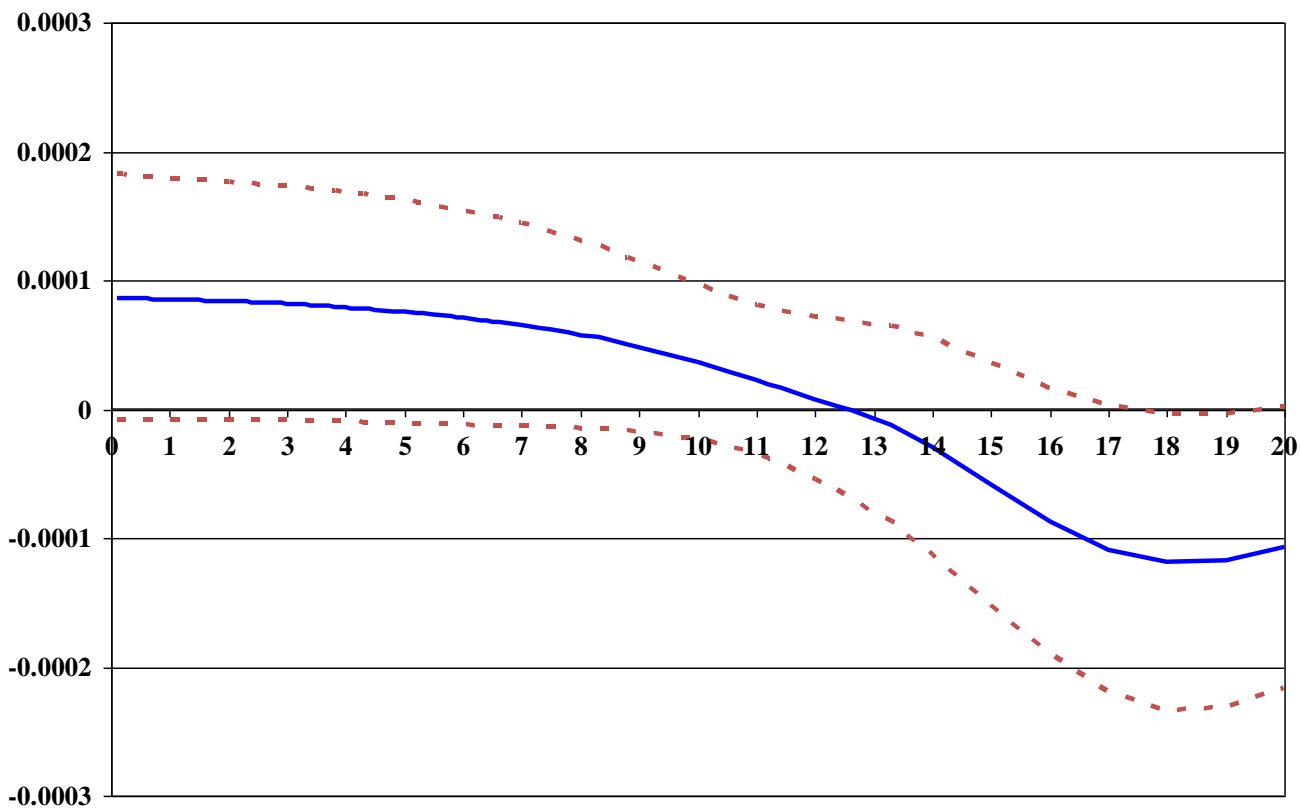
Figure 10. Cost per Horsepower for Electric Motors and Submersible Pumps by Horsepower (ICRISAT VLS Equipment Inventory, July 2011)



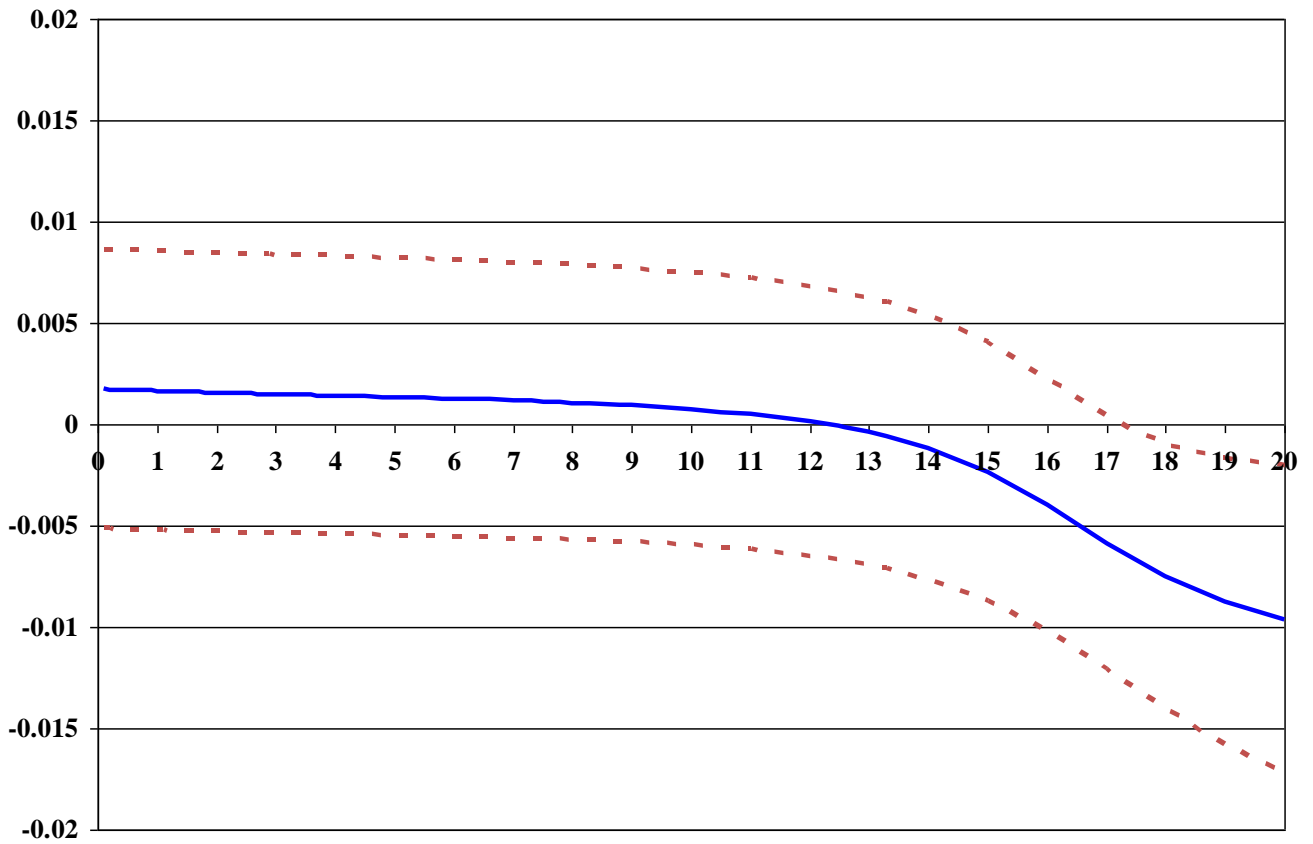
**Figure 11. Average Hourly Wage Paid for Male Labor, by Farm Size
(ICRISAT VLS 2009-14)**



**Figure 12. LWFCM Plot Fixed Effect Estimates:
The Effect of Rainfall on the Fraction of Operations Using Low-Hours Hired Male Labor,
with 95% CI, by Plot Size (ICRISAT VLS 2009-14)**



**Figure 13. LWFCM Plot Fixed-Effect Estimates:
the Effect of Rainfall on the Average Male Wage, with 95% CI, by Plot Size
(ICRISAT 2009-14)**



**Figure 14. Fraction of Farms using Sprayers and Tractors, by Farm Size
(ICRISAT VLS 2009-14)**

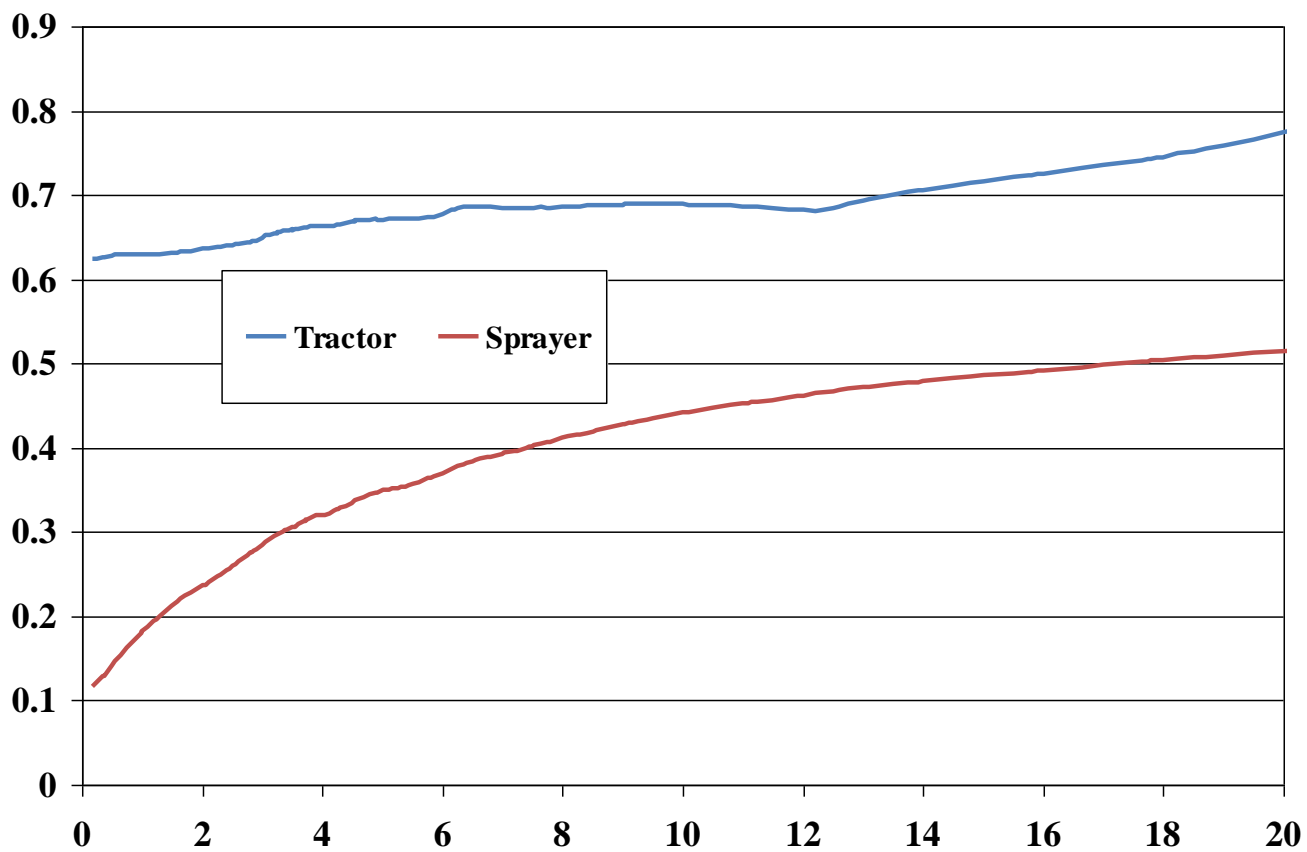


Figure 15. Per-Acre Equipment Hours for Tractors and Sprayers, by Farm Size (ICRISAT VLS 2009-14)

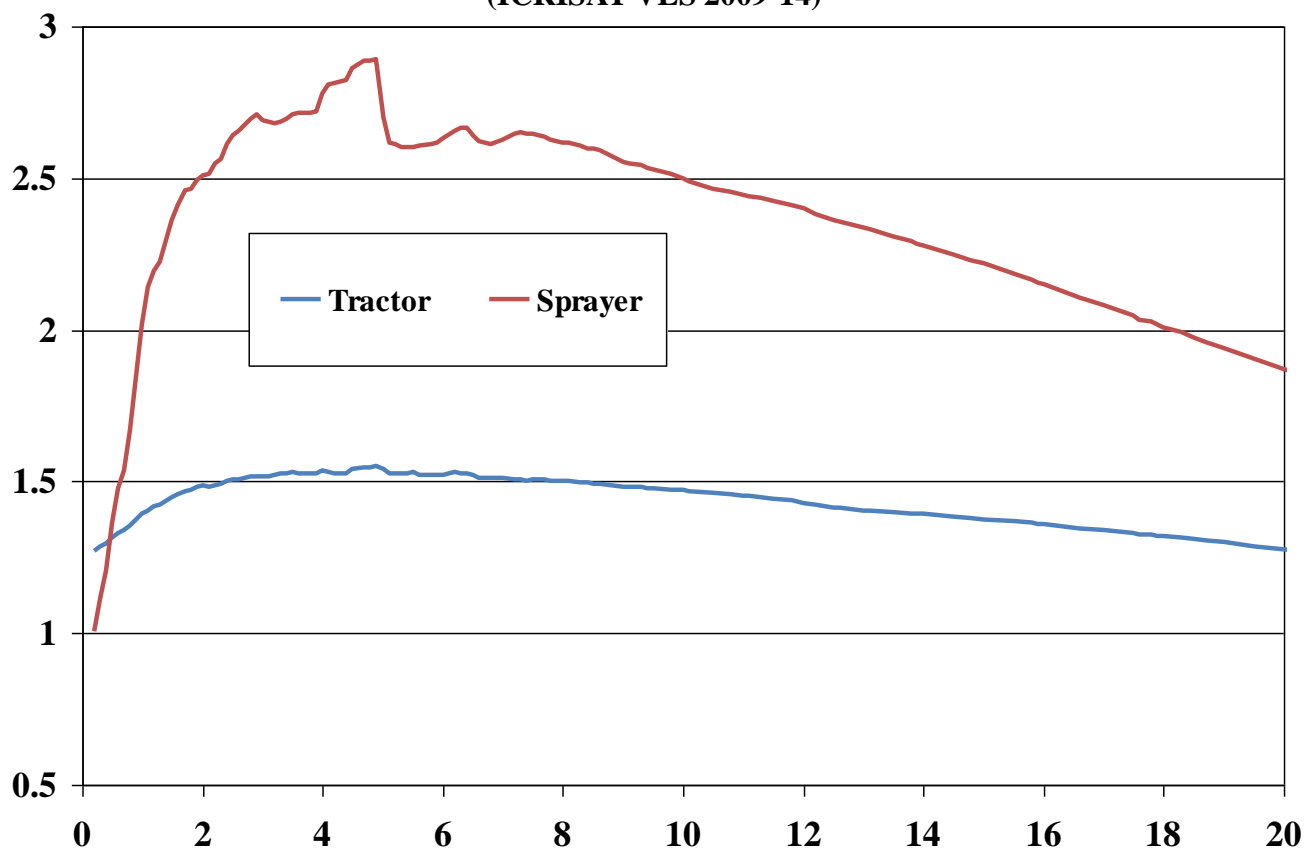


Figure 16. Real Weeding and Sprayer Labor Costs and Total Sprayer Costs per Acre, by Farm Size (ICRISAT VLS 2009-14)

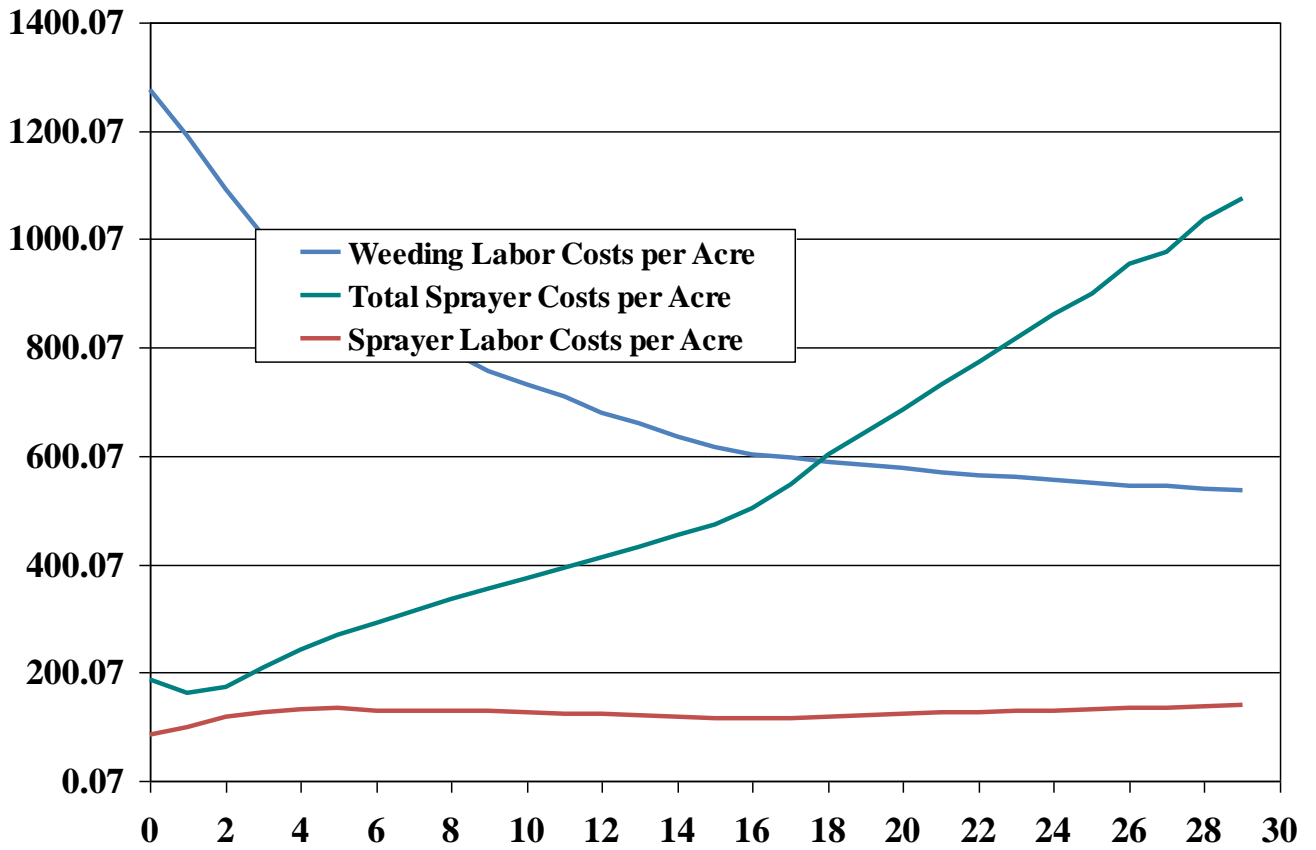


Figure 17: Model Fit: Profit per Acre and Land Size

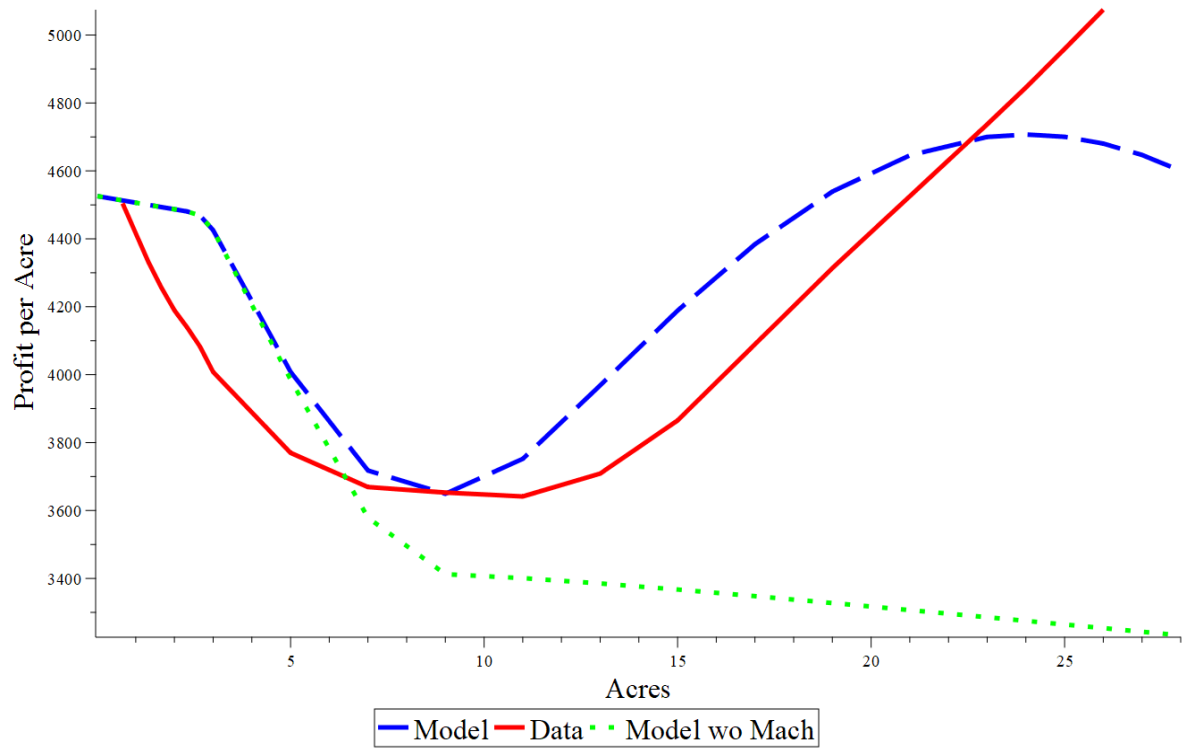


Figure 18: Model Fit: Output per Worker and Land Size

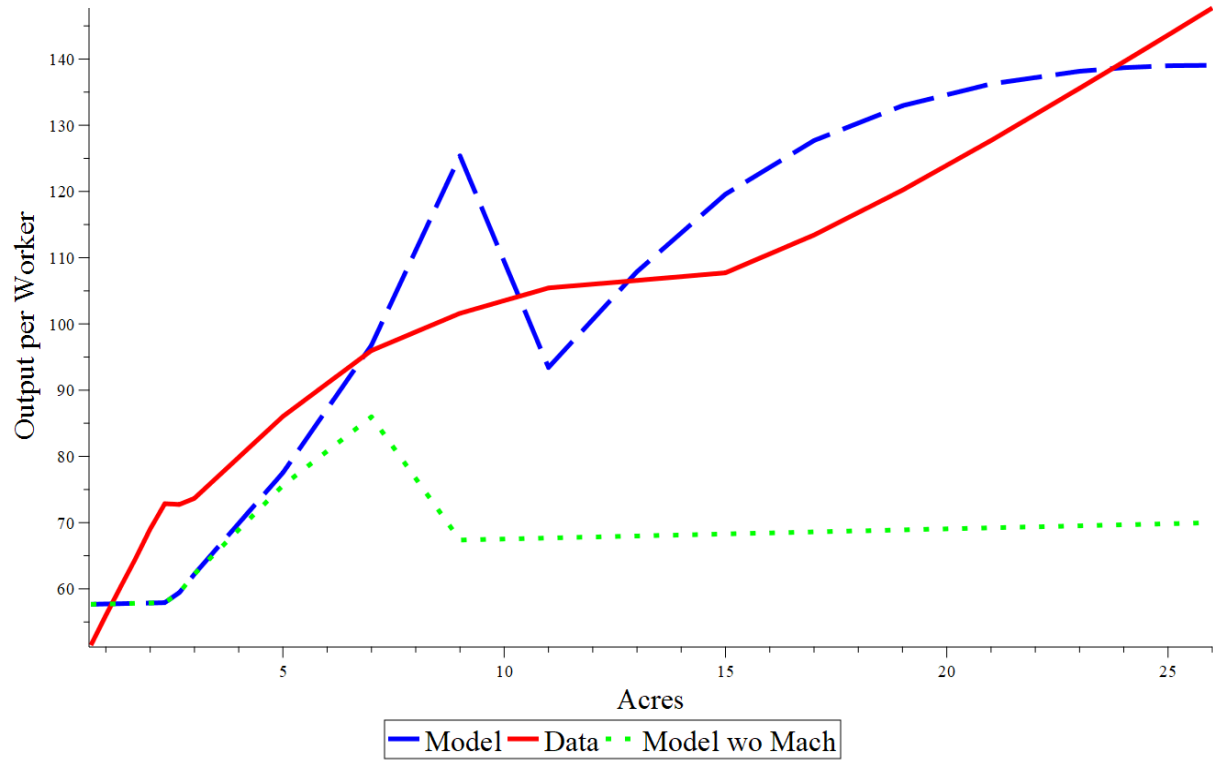
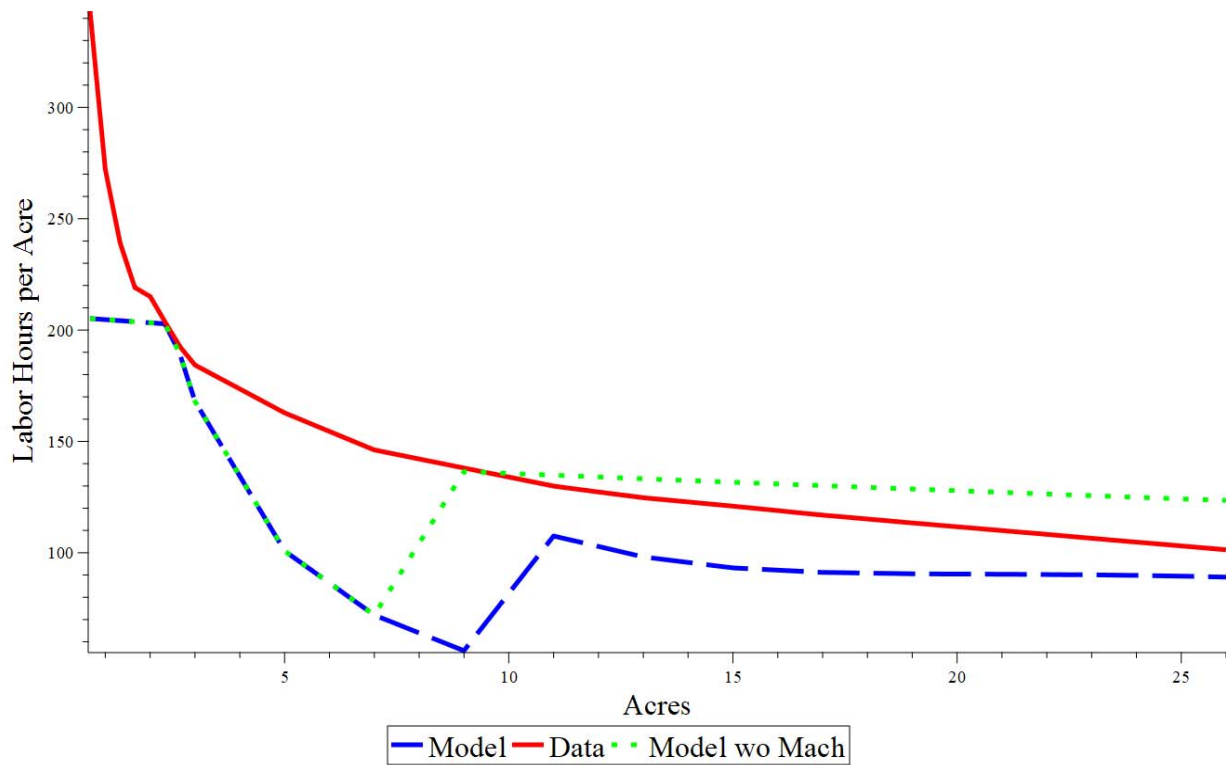
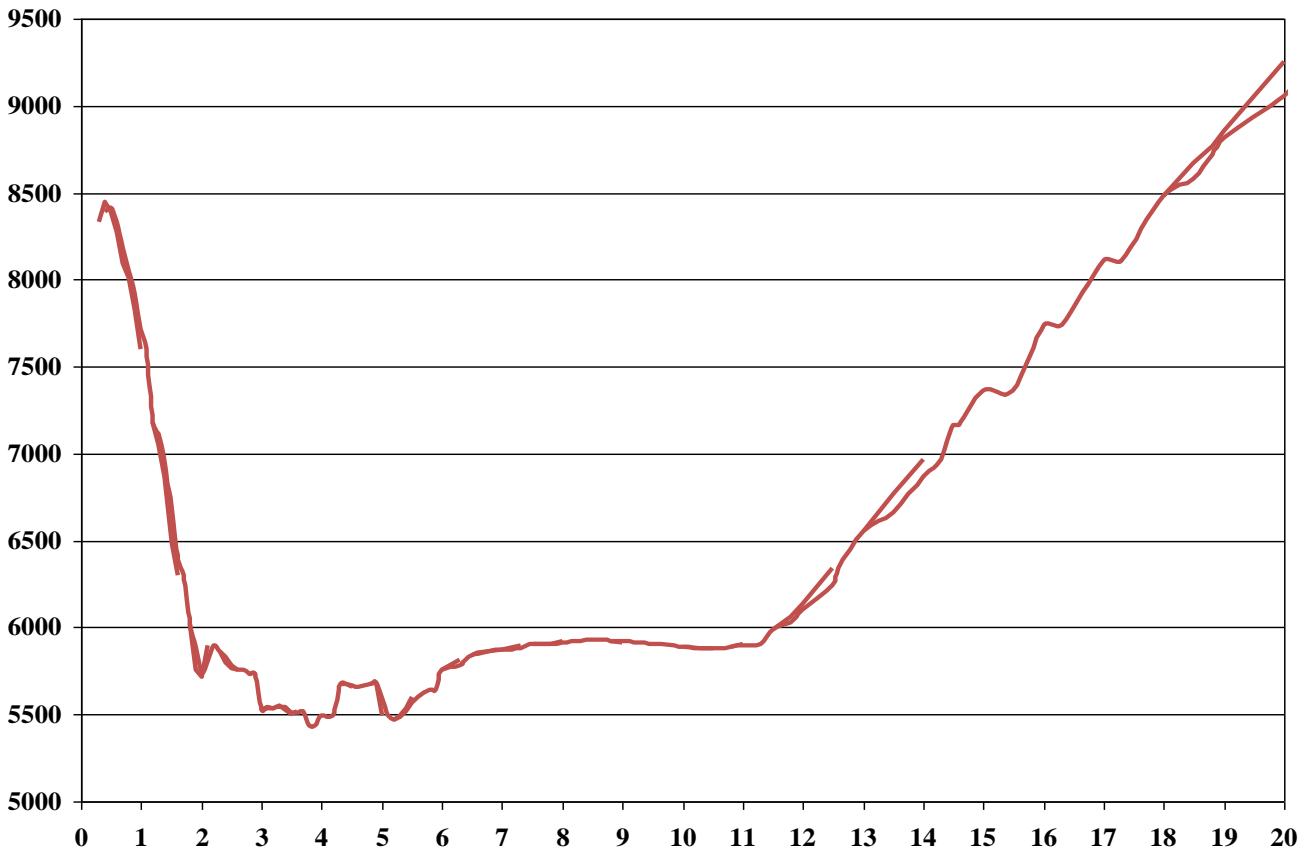


Figure 19: Model Fit: Labor Hours per Acre and Land Size



**Appendix Figure A1. Lowess-Smoothed Relationship of Profits per Acre and Owned Plot Size,
Cotton Plots Only
(ICRISAT Survey, 2009-14)**



**Appendix Figure A2. LWFCM Measurement Error Effects: Profits and Owned Acreage
Coefficient Point Estimates, by Farm Acreage and Estimation Procedure
(ICRISAT VLS 2009)**

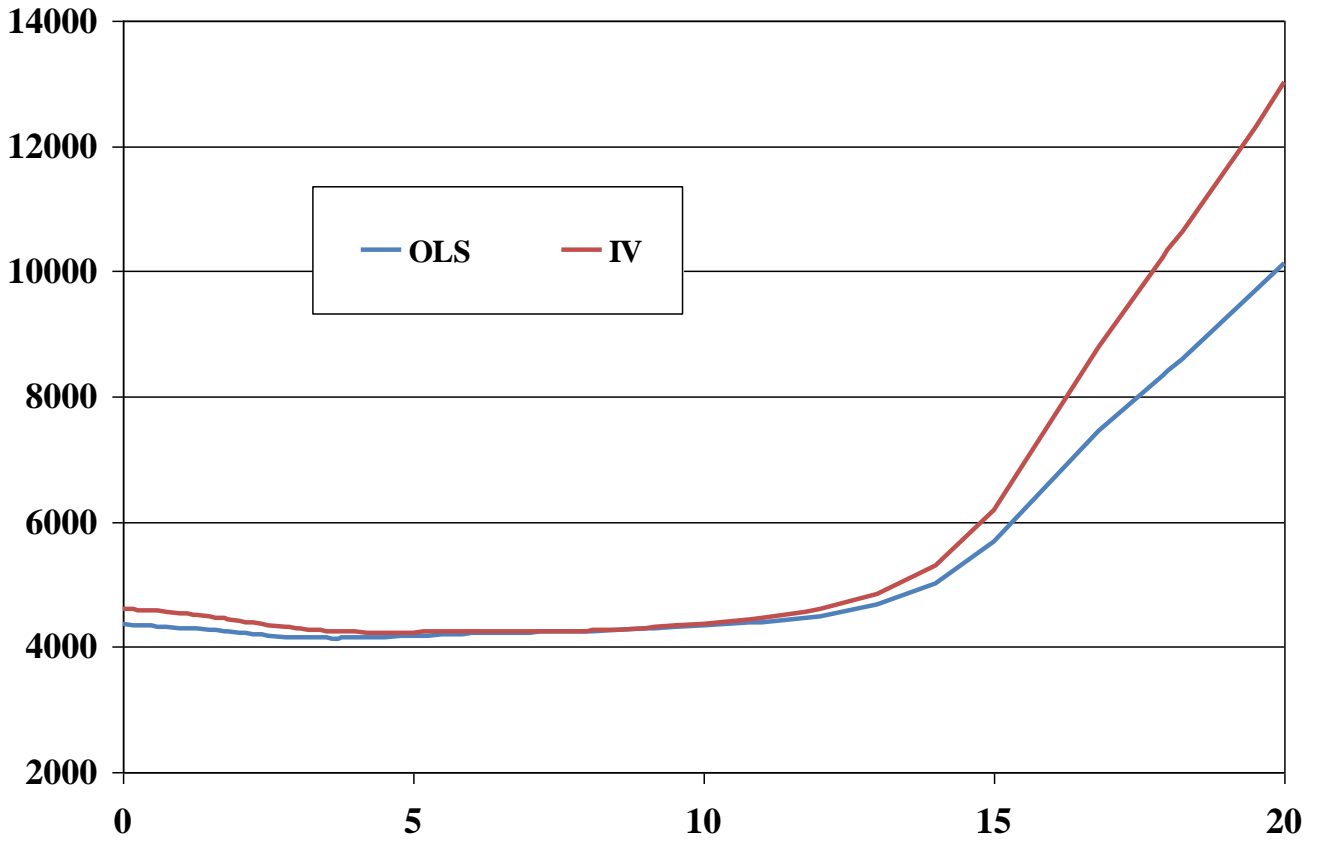


Table 1
 Operation Fixed Effects Estimates: Percent Difference in Hourly Wage Rates Paid
 for Eight Hours versus Less than Eight Hours of Work, by Input and Data Source

Input	Hired Male Labor		Hired Bullock Pair + Driver		Sprayer
Data source	2010, 2011 Monthly Price Schedules	2014 Input Survey	2010, 2011 Monthly Price Schedules	2014 Input Survey	2009-2014 Input Surveys ^b
Worked eight hours in the day versus <8	-33.2 (3.14)	-34.7 ^a (11.9)	-22.3 (4.54)	-30.0 ^a (8.42)	-13.2 ^a (13.1)
Log capacity	-	-	-	-	0.626 ^a (0.128)
Mean wage (rupees)	22.1 (9.34)	34.7 (19.3)	78.7 (39.6)	114.2 (35.0)	15.9 (23.3)
Percent working <8 hours	30.7	44.4	58.4	61.0	19.3
N	729	3,387	450	1,240	1,201

^aFE-IV estimate; first-stage includes log of owned area and all land quality characteristics.

^bSpecification also includes village-year fixed-effects.

Standard errors clustered at the village-year level in parentheses.

Hourly wage rate = daily wage/hours worked. Sprayer capacity = material sprayed per hour of use.

Table 2
 FE-IV Estimates: the Percentage Hourly Cost Discount by Plot Area and Plot Distance from the Homestead,
 for Male Hired Workers and Rented Sprayers^a
 Dependent Variable=Log Hourly Wage/Rental Price

Input	Hired Male Labor (2014)		Sprayer (2009-2014)	
Worked eight hours in the day versus <8	-12.6 (12.7)	-6.21 (12.6)	-7.59 (14.4)	-2.02 (2.20)
Worked eight hours in the day*plot area	-31.2 (6.84)	-32.0 (6.98)	-4.86 (4.91)	-6.28 (5.04)
Worked eight hours in the day*plot distance from the homestead	-	-13.5 (2.78)	-	-11.1 (12.1)
Log capacity	-	-	0.582 (0.135)	0.423 (0.139)
Operation fixed effects	Y	Y	N	N
Village-year fixed effects	N	N	Y	Y
N	3,387		1,201	

^aFirst-stage includes owned area and all land quality characteristics.
 Standard errors in parentheses clustered at the village-year level.
 Sprayer capacity = material sprayed per hour of use.

Table 3
 Plot Size and Fraction of Operations that Employ Hired Inputs at Low (≤ 6) Daily Hours and the Average Hourly Wage Paid,
 by Input Type (*Kharif* Seasons 2009-14)

Variable	Fraction of Operations ≤ 6 Hours/Day			Average Hourly Wage		
	Hired Male Labor	Hired Tractor	Hired Bullock Pair	Hired Male Labor	Hired Tractor	Hired Bullock Pair
Plot size (acres)	-.0165 (.00306)	-.0197 (.00247)	-.0170 (.00306)	-.183 (.0876)	1.25 (.769)	-.866 (.306)
Plot size squared $\times 10^{-3}$.450 (.112)	.449 (.0682)	.555 (.117)	8.29 (3.23)	18.3 (32.4)	29.3 (10.9)
Village/year FE	Y	Y	Y	Y	Y	Y
25 Plot and household characteristics	Y	Y	Y	Y	Y	Y
Number of observations	6,777	6,777	6,777	6,777	6,777	6,777

Standard errors in parentheses clustered at the village/year level.

Table 4
 Plot Fixed Effects Estimates: The Effects of *Kharif*-Season Rainfall on Profits, Hours Employed and Average Hourly Wage Rates,
 by Input Type (*Kharif* Seasons, 2009-14)

Variable	Profits	Hours Employed			Average Hourly Wage		
Input type	-	Hired Male Labor	Hired Tractor	Hired Bullock Pair	Hired Male Labor	Hired Tractor	Hired Bullock Pair
Rainfall (mm)	38.1 (17.1)	.182 (.0701)	.00362 (.00316)	.0347 (.0248)	-.0158 (.00672)	.0130 (.0601)	-.0593 (.0355)
Rainfall squared x10 ⁻³	-21.2 (8.59)	-.107 (.0377)	-.00214 (.00161)	-.0500 (.0268)	.00778 (.00398)	-.0132 (.0282)	.0757 (.0331)
Year and plot FE	Y	Y	Y	Y	Y	Y	Y
H ₀ : Rain and rain squared = 0 F(2,n) [p]	3.09 [.0504]	4.18 [.0183]	0.99 [.3742]	1.97 [.1452]	3.47 [.0352]	0.28 [.7589]	3.02 [.0538]
Number of observations	5,291	3,987	4,016	2,523	3,987	4,016	2,523

Standard errors in parentheses clustered at the village/year level.

Table 5
Farm Size, Wealth and Mechanization (Ownership): 2014 ICRISAT VLS Round

Variable	Owns a Tractor	Owns a Power Sprayer	
Sample	All Farmers	All Farmers	Farmers Who Own Any Sprayer
Total owned land (acres)	.0125 (.00415)	.0107 (.00474)	.0133 (.00494)
Total rental value of land (wealth) x 10 ⁻⁵	.0506 (.0146)	.0512 (.0166)	.0273 (.0144)
Village FE	Y	Y	Y
Percent owning	3.5	10.3	24.8
Number of farmers	652	652	288

Standard errors in parentheses clustered at the village level. All specifications include the head's age and schooling.

Table 6
Cost and Capacities of Indian KrisanKraft Power Sprayers, 2017





Power sprayer	Litres/Hour	Current Price (Rupees)
 <p>KK-708 Knapsack Power Sprayer</p>	180	7830
 <p>KK-PPS-P764 Portable Power Sprayer</p>	420	12260
	1320	25900
	2400	27900

Table 7
 Estimates of the Effects of Owned Land Size
 on Sprayer Use, Weeding Hours per Acre, Sprayer Hours per Acre, Log Sprayer Price per hour, and Sprayer Flow Rate

Variable	Any sprayer use	Weeding hours per acre	Sprayer hours per acre	Sprayer log price per hour	Sprayer flow rate
Estimation procedure	OLS	OLS	OLS	OLS	OLS
Owned area	0.006197 (0.0009879)	-0.5631 (0.1286)	-0.4063 (0.0853)	0.01335 (0.00669)	0.01360 (0.00667)
All land characteristics	Y	Y	Y	Y	Y
Village/year fixed effects	Y	Y	Y	Y	Y
N	3,374	3,374	1,219	1,219	1,219

Standard errors in parentheses clustered at the village/year level.

Table 8
GMM Estimates of the Effective Capacity Function $\varphi(a)$ and Price Parameter ν

Coefficient	Point Estimate	Robust SE
ν	0.316	0.124
b_0	5.58	0.0375
b_1	0.933	0.0343
b_2	-0.0190	0.00211
$H_0: \nu < 1, \chi^2(1)$ [p]	30.4 [.0000]	
Maximum land size (acres) = $\varphi(a)' = -b_1/(2*b_2) = 0$	24.5	1.84
N	617	

Instruments: owned land area and land area squared.

Table 9
Estimates of Sprayer ν , by Source

Country	India		United States
Source	ICRISAT Survey, 2009-2014	KrisanKraft Price List (2016)	Stiles and Stark (2016)
Estimation procedure	IV ^a	OLS	OLS
ν	0.5802 (0.1200)	0.5209 (0.0605)	0.1458 (0.0789)
$H_0: \nu = 1, F(1,x)$ [p]	$\chi^2=12.2$ [0.0005]	$F(1,2)= 62.8$ [0.0156]	$F(1,2)=117.1$ [0.0084]
N	1,219	4	4
Village/year fixed effects	Y	N	N

^aFirst-stage includes log of owned area and all land quality characteristics. Standard error clustered at the village/year level.

Table 10
Calibrated Equilibrium Effects of Making All Farms of Optimal Size

Scenario	Baseline	Post-Reform	
Urban wage elasticity	-	0	-0.4
Average farm size (acres)	3.13	24.0	24.1
Number of farms (millions)	95.2	12.4	12.4
Profits per acre (Rs)	4276	4705	4845
Total agricultural output (trillion Rs)	2.71	3.71	3.85
Profit per worker (Rs)	6302	9634	9225
Output per worker (thousand Rs)	14.5	25.4	24.4
Hourly wage (Rs)	21	21	19.2
Size of agricultural labor force (million)	187	146	158
Work hours per farm	361	2157	2340
Machine hours per farm	58.5	163	162
Fraction of farms using machines	.213	1	1
Machine capacity index (mode)	4.49	7.74	7.72

Appendix Table A1
Measurement Error Test: The Effect of Total Owned Area on Profits,
by Estimation Procedure, Using the Survey and Census Owned Land Self-reports in 2009

Variable	Profits		Owned Land - Survey
	OLS	IV	OLS
Total owned land - survey, sum of plot sizes	4324.2 (1305.8)	4433.5 (1328.1)	-
Owned land - Census, sum of irrigated and non-irrigated land	-	-	0.9637 (0.0128)
Village fixed effects	Y	Y	Y
N	306	306	306

Standard errors in parentheses clustered at the village level. Profits are for the *Kharif* season.

Appendix Table A2
Does Land Quality Affect Estimates of the Relationship Between Real Output Value and Profits and Area?
(*Kharif* Seasons 2009-14)

Variable	Real Output Value				Real Profits			
Farm size (acres)	9138.9 (676.3)	9053.4 (671.2)	-	-	3911.8 (501.4)	3871.9 (500.7)	-	-
Plot size (acres)	-	-	7390.6 (1463.8)	7548.6 (1506.0)	-	-	5841.8 (1474.0)	5939.5 (1521.7)
Village/year FE	Y	Y	-	-	N	Y	-	-
24 plot characteristics/categories	N	Y	N	Y	N	N	N	Y
Farmer/year fixed effects	N	N	Y	Y	N	N	Y	Y
H ₀ : Plot and household characteristics = 0 F(24,107) [p]	-	8.45 [.0000]	-	1.22 [.239]	-	3.61 [.0000]	-	1.22 [.245]
Number of observations	3,835	3,835	7,865	7,865	3,835	3,835	7,865	7,865

Standard errors in parentheses clustered at the village/year level.

Appendix Table A3
Mean Standard Deviations (Days) of First Operation Dates Across Plots
Within and Across Farmers, by Operation
(*Kharif* Season 2014)

Operation/sample	Cross-Plot within Farmer ^a	Cross-Farmer within Village
Sowing	9.55 (14.7) N=713	14.8 (10.9) N=946
Fertilizing	20.2 (19.7) N=464	24.8 (14.2) N=647
Weeding	16.6 (15.9) N=737	19.3 (9.8) N=984
Spraying	17.7 (19.5) N=447	22.4 (12.0) N=664
Threshing	13.7 (13.5) N=468	18.0 (8.31) (678)
Harvesting	21.8 (19.7) N=914	24.9 (12.4) N=1,133

^aFarmers with two or more plots. Standard errors in parentheses.